STATISTICAL ANALYSIS PLAN

Title: A Multicenter, International, Randomized, Double-blind, Alendronate-controlled Study to Determine the Efficacy and Safety of Romosozumab in the Treatment of Postmenopausal Women With Osteoporosis

Protocol Number: **20110142**

Version: **3.0**

Date: **07 April 2017**

Author: PPD **and** PPD

NCT Number: 1631214
This NCT number has been applied to the document
for purposes of posting on clinicaltrials.gov

Table of Contents

List of Tables

Table of Abbreviations

Abbreviation	Definition
AB	Antibody
AE	Adverse Event
ALN	Alendronate
ANCOVA	Analysis of Covariance
BMD	Bone Mineral Density
BMI	Body Mass Index
BPI	Brief Pain Index
BTM	Bone Turnover Marker
BSAP	Bone Specific Alkaline Phosphatase
ClinRO	Clinician Reported Outcome; questionnaires completed by physician or site staff using information provided by the subject
CPEVENT	Clinical planned event
CTCAE	Common Terminology Criteria of Adverse Events
CT	Computerized Tomography
CV	Cardiovascular
DMC	Data Monitoring Committee
DXA	Dual X-ray Absorptiometry
eCRF	Electronic Case Report Form
eGFR	Estimated Glomerular Filtration Rate
EDC	Electronic Data Capture
EQ-5D-5L	EuroQol-5 Dimensions-5 Levels Health Survey
FEA	Finite Element Analysis
FRAX	10-year Probability of Major Osteoporotic and Hip Fractures Based on WHO Risk Factor Criteria
ICF	Informed Consent Form
IP	Investigational Product
IPD	Important Protocol Deviation
iPTH	Intact Parathyroid Hormone
IV	Intravenous
IVRS	Interactive Voice Response System
LAD	Limited Activity Days

Abbreviation	Definition
LOCF	Last Observation Carried Forward
MDRD	Modification of Diet in Renal Disease
MedDRA	Medical Dictionary for Regulatory Activities
MRI	Magnetic Resonance Imaging
OPAQ SV	Osteoporosis Assessment Questionnaire Short Version
OC	Osteocalcin
P1NP	Procollagen Type I N-terminal Peptide
PK	Pharmacokinetic
PKDM	Pharmacokinetic and Drug Metabolism
PO	Orally
PRO	Patient Reported Outcomes
PTH	Parathyroid Hormone
QCT	Quantitative Computerized Tomography
QM	Every month
sCTx	Serum Type I Collagen C-telopeptide
SAP	Statistical Analysis Plan
SC	Subcutaneous
SERM	Selective Estrogen Receptor Modulator
SMQ	Standardised MedDRA Query
VAS	Visual Analog Scale
vBMD	Volumetric Bone Mineral Density
WHO	World Health Organization

1. Introduction

The purpose of this Statistical Analysis Plan (SAP) is to provide details of the statistical analyses that have been outlined within the protocol for romosozumab, Protocol 20110142 dated 14 September 2016. The scope of this plan includes analyses that are planned and will be executed by the Biostatistics department or designee unless otherwise specified (eg, pharmacokinetic [PK] modeling will be provided by Pharmacokinetic and Drug Metabolism [PKDM] group).

2. Objectives

2.1 Primary

For the primary analysis period (randomization to primary analysis):

To assess the effect of romosozumab treatment for 12 months followed by alendronate (ALN) treatment compared with ALN treatment alone on:

- Subject incidence of clinical fracture (nonvertebral fracture and clinical vertebral fracture) in postmenopausal women with osteoporosis
- Subject incidence of new vertebral fracture in postmenopausal women with osteoporosis

2.2 Secondary

For the primary analysis period:

To assess the effect of romosozumab treatment for 12 months followed by ALN treatment compared with ALN treatment alone on:

- Subject incidence of fractures (all fractures [nonvertebral fractures and new or worsening vertebral fractures], new or worsening vertebral fracture, nonvertebral fracture, major nonvertebral fracture [pelvis, distal femur, proximal tibia, ribs, proximal humerus, forearm, and hip], hip fracture, multiple new or worsening vertebral fracture, and clinical vertebral fracture)
- Percent changes in Dual energy X-ray Absorptiometry (DXA) bone mineral density (BMD) at the lumbar spine, total hip, and femoral neck

For the 12-month double-blind ALN-controlled study period:

To assess the effect of romosozumab treatment for 12 months compared with ALN treatment on:

- Subject incidence of fractures (clinical fracture [nonvertebral fracture and clinical vertebral fracture], new vertebral fracture, all fractures [nonvertebral fractures and new or worsening vertebral fractures], nonvertebral fracture, hip fracture, clinical vertebral fracture, major osteoporotic fracture [hip, forearm, humerus, and clinical vertebral])

- Percent changes from baseline in DXA BMD at the lumbar spine, total hip and femoral neck

For the overall study (randomization to end of study):

- To assess the effect of romosozumab treatment for 12 months followed by ALN treatment compared with ALN treatment alone on subject incidence of hip fracture, major nonvertebral fracture [pelvis, distal femur, proximal tibia, ribs, proximal humerus, forearm, and hip], and nonvertebral fractures

2.3 Exploratory

For the 12-month double-blind ALN-controlled study period:

- To assess the effect of romosozumab treatment for 12 months compared with ALN treatment on subject incidence of fractures (new or worsening vertebral fracture, major nonvertebral fracture [pelvis, distal femur, proximal tibia, ribs, proximal humerus, forearm, and hip], and multiple new or worsening vertebral fracture)
- To describe subjects' experience of pain and its impact on subjects' activities and health-related quality of life at 6-month intervals for postmenopausal women with osteoporosis using Patient Reported Outcome (PRO) and Clinician Reported Outcome (ClinRO) questionnaires (Osteoporosis Assessment Questionnaire Short Version [OPAQ SV], EuroQoL-5 Dimensions-5 Levels Health Survey [EQ-5D-5L], Limited Activity Days [LAD] survey, and one item extracted from the Brief Pain Inventory [BPI] assessing the worst pain experienced in the past 24 hours [BPI worst pain])
- To describe subjects' experience of pain and its impact on subjects' activities and health-related quality of life for 3 months at 1-month intervals after experiencing a nonvertebral or clinical vertebral fracture (OPAQ SV, EQ-5D-5L, LAD, and BPI worst pain)
- To describe the effect of romosozumab treatment for 12 months compared with ALN treatment on the percent of subjects with a clinically meaningful improvement in worst pain for 3 months at 1-month intervals after experiencing a nonvertebral or clinical vertebral fracture (defined as a 2-point improvement in the BPI worst pain scale compared with the fracture reporting visit)
- To assess the effect of romosozumab treatment for 12 months compared with ALN treatment on changes in height

For the primary analysis period:

- To assess the effect of romosozumab treatment for 12 months followed by ALN treatment compared with ALN treatment alone on subject incidence of major osteoporotic fracture (hip, forearm, humerus, and clinical vertebral)
- To describe subjects' experience of pain and its impact on subjects' activities and health-related quality of life at 6-month intervals for postmenopausal women with osteoporosis using PRO and ClinRO questionnaires (OPAQ SV, EQ-5D-5L, LAD, and BPI worst pain)
- To assess the effect of romosozumab treatment for 12 months followed by ALN treatment compared with ALN treatment alone on changes in height

During the Month 12 to 24 ALN study period

- To assess the effect of one year of romosozumab treatment compared with ALN treatment alone on subject incidence of new vertebral fractures, clinical fracture (nonvertebral fracture and clinical vertebral fracture), nonvertebral fracture, hip fracture, and clinical vertebral fracture during the subsequent year during which all subjects receive ALN treatment.

2.4 Safety

For the 12-month double-blind ALN-controlled study period:

- To characterize the safety and tolerability of romosozumab treatment for 12 months as determined by a review of reported adverse events, laboratory data, vital signs, and formation of anti-romosozumab antibodies

For primary analysis period:

- To characterize the safety and tolerability of romosozumab treatment for 12 months followed by ALN treatment as determined by a review of reported adverse events, laboratory data, vital signs, and formation of anti-romosozumab antibodies

For the overall study:

- To characterize the safety and tolerability of romosozumab treatment for 12 months followed by ALN treatment as determined by a review of reported adverse events and vital signs

2.5 Imaging and Pharmacokinetics (PK) / Bone Turnover Marker (BTM) / Biomarker Sub-study

For the 12-month double-blind ALN-controlled study period:

- To characterize the serum romosozumab concentration
- To assess the effect of romosozumab treatment for 12 months compared with ALN treatment on:
  - Percent changes in bone formation markers Procollagen Type 1 N-telopeptide (P1NP), Bone Specific Alkaline Phosphatase (BSAP), and Osteocalcin (OC), and in bone resorption marker serum Type I collagen C-telopeptide (sCTX)
  - Percent changes in sclerostin and intact Parathyroid Hormone (iPTH)
  - For subjects also participating in the imaging components: Percent changes in integral (total) and trabecular volumetric BMD (vBMD) at the lumbar spine by Quantitative Computed Tomography (QCT)
  - For subjects also participating in the imaging components: Percent changes in lumbar spine strength as assessed by Finite Element Analysis (FEA)
- For subjects participating in the imaging components: To assess the effect of romosozumab treatment for 6 months compared with ALN treatment for 6 months on percent changes in DXA BMD at the lumbar spine, total hip, and femoral neck

- To enable exploratory assessments of novel biomarkers through prospective collection of blood samples

For the primary analysis period:

- To assess the effect of romosozumab treatment for 12 months followed by ALN treatment compared with ALN treatment alone on:
  - Percent changes in bone formation marker P1NP and bone resorption marker sCTX
  - Percent changes in sclerostin and iPTH
  - For subjects participating in the imaging components: Percent changes in integral (total) and trabecular vBMD at the lumbar spine by QCT
  - For subjects participating in the imaging components: Percent changes in lumbar spine strength as assessed by FEA
- For subjects participating in the imaging components: To assess the effect of romosozumab treatment for 12 months followed by ALN treatment for 6 months compared with ALN treatment alone on percent changes in DXA BMD at the lumbar spine, total hip, and femoral neck.
- To enable exploratory assessments of novel biomarkers through prospective collection of blood samples

3. Study overview

3.1 Study Design

This is a phase 3 multicenter, international, randomized, double-blind, ALN-controlled study of romosozumab in postmenopausal women with osteoporosis. The study is designed to evaluate if romosozumab treatment for 12 months followed by ALN treatment, compared with ALN treatment alone, is effective in reducing the incidence of clinical fracture (nonvertebral fracture and clinical vertebral fracture) and/or new vertebral fracture.

After signing the informed consent form (ICF), subjects will undergo a screening phase to complete eligibility assessments. Upon confirmation of eligibility, approximately 4,000 subjects will be randomized 1:1 (approximately 2,000 subjects per arm) to receive either 210 mg romosozumab subcutaneous (SC) every month (QM) or 70 mg ALN orally (PO) every week (QW) in a blinded fashion for the duration of the 12–month double-blind ALN-controlled study period. Randomization will be stratified by age ($<$ 75 years, $\geq$ 75 years). Subjects will also receive matched placebo for either ALN or romosozumab. After the initial 12-month study period, subjects will receive ALN while

remaining blinded to their initial treatment assignment (romosozumab or ALN). The primary analysis period will end and the primary analysis will be performed when:

- clinical fracture events (nonvertebral fracture or clinical vertebral fracture) have been confirmed for at least 330 subjects
  AND
- all subjects have had the opportunity to complete their Month 24 study visit.

If more than 330 subjects have confirmed clinical fractures at the time each subject has completed her Month 24 visit, the primary analysis will be based on all available data, which may include more than 330 events.

Upon completion of the primary analysis period, subjects will continue to be followed for the secondary endpoint of nonvertebral fractures. Subjects and site personnel will remain blinded to initial treatment assignments. The study will proceed in an event-driven manner. The final analysis (end of study) will be performed when nonvertebral fracture events have been confirmed for at least 440 subjects across the lifetime of the study. The study may be terminated earlier if the primary analysis demonstrates superiority of romosozumab treatment for nonvertebral fracture risk reduction (see Section 3.2.3 for details on testing strategy). Study completion constitutes either completion of primary analysis if superiority is proven for nonvertebral fractures or until 440 subjects experience a nonvertebral fracture.

Approximately 200 subjects at participating centers will be enrolled in an Imaging and PK/BTM/Biomarker sub-study. Within this sub-study, a subset of approximately 100 subjects will participate in the imaging (DXA, QCT) portion of the sub-study.

If there is insufficient enrollment into the blood only portion of the sub-study, additional sub-study sites will be intiatied. Subjects already on trial will be invited to participate and consent will be obtained for the use of blood samples from earlier time points for sub-study analyses.

From study start until primary analysis, an external, independent Data Monitoring Committee (DMC) will monitor unblinded safety data on an ongoing basis throughout the study and, if a potential safety signal is identified, may consider efficacy data in order to assess the risk/benefit profile of romosozumab.

3.2 Sample Size

3.2.1 Sample Size Considerations

The study is designed to evaluate the effectiveness of romosozumab treatment for 12 months followed by ALN (romosozumab /ALN) treatment compared with ALN treatment alone in reducing the subject incidence of clinical fracture (nonvertebral fracture or clinical vertebral fracture), subject incidence of vertebral fracture through Month 24, and subject incidence of nonvertebral fracture in postmenopausal women with osteoporosis. The total sample size of approximately 4,000 subjects (2,000:2,000 equally allocated between the treatment groups) is determined based on the clinical fracture and new vertebral fracture endpoints.

The dropout rate is assumed 10% for the first year and 8% per year thereafter. The enrollment is expected to complete in 34 months. Because study design requires all subjects to complete the Month 24 visit for the ascertainment of the new vertebral fracture endpoint, the minimum follow-up on an individual subject is at least 24 months. Additionally, all subjects will be followed until 440 subjects have confirmed nonvertebral fracture events for the final analysis, unless superiority of the nonvertebral fracture endpoint is achieved at the primary analysis. The minimum follow-up time (24 months) and event-driven components of this study add complexity to the sample size calculation and power estimates. To address this complexity and the added complexity of using Hochberg's method (Hochberg, 1988) to adjust for the multiplicity of the primary endpoints, a simulation was performed to derive the sample size and power for the clinical fracture and nonvertebral fracture endpoints.

3.2.2 Assumptions

New vertebral fracture assumptions

The one-year incidence of new vertebral fractures in untreated subjects in the population defined in this study is expected to be 5.5% (Black et al, 1996; Cummings et al, 2009). Assuming ALN treatment decreases this incidence by 60%, the new vertebral fracture incidence in the ALN group is expected to be 2.2% in one year and 4.4% in two years. **romosozumab** /ALN treatment is expected to decrease the incidence of new vertebral fractures by 80%. This represents a risk reduction of 50% compared to ALN treatment alone.

Nonvertebral fracture assumptions

The one-year incidence rate of nonvertebral in the population defined in this study is expected to be 5.5% (Black et al, 2000). Based on the ALN treatment assumption, 25% risk reduction, the one-year nonvertebral incidence rate is expected to be approximately 4% when treated by ALN. Romosozumab /ALN is assumed to reduce the risk by 45% when compared to no treatment. The combination of the 25% assumed reduction for ALN and the 45% assumed reduction for romosozumab /ALN leads to the assumption of a 27% reduction in fracture risk when comparing romosozumab /ALN to ALN.

Clinical fracture assumptions

Clinical fracture is the combination of nonvertebral fractures and clinical vertebral fractures. Clinical vertebral fractures are expected to account for 1/3 (Black et al, 2000) of new vertebral fractures. In the FREEDOM study, in the untreated population, the one-year incidence rate was 0.5% for clinical vertebral fracture and 3.5% for clinical fracture (Amgen, data on file). Therefore, it is expected that clinical fractures are composed of approximately 85% nonvertebral fractures and 15% clinical vertebral fractures. By pooling the treatment effects of nonvertebral and clinical vertebral fractures in the natural log-scale, with 85% and 15% as weights, then taking the exponential, romosozumab /ALN treatment compared with ALN alone reduces the risk of clinical fractures by approximately 30% (ie, exp(0.15*ln(1-50%) + 0.85*ln(1-27%)) = 69%, approximately 30% risk reduction). This estimation of risk reduction was further supported in the simulation. Under the assumption that romosozumab /ALN will reduce the risk by 30% compared to ALN, we would need to follow subjects until the 330th subject had a confirmed clinical fracture to achieve a 90% power to detect the treatment effect using a 2-sided log-rank test at an overall significance level of 0.05. The power calculation was performed using software EAST 5.4.

DXA BMD assumptions

The assumptions of mean percent change in DXA BMD and standard deviation for ALN at Month 12 and Month 24 are summarized in Table 1 (Schnitzer et al, 2000; Greenspan et al, 2002). Using the FRAME study, the assumptions of mean percent change in DXA BMD and standard deviation for romosozumab at Month 12 are

summarized in Table 2 (Amgen, data on file). The mean of percent change from baseline at hip trochanter is used as the assumption for total hip. It is further assumed that at Month 24, romosozumab/ALN will maintain the BMD increase achieved at Month 12.

Table 1. Mean Percent Change From Baseline in BMD for Alendronate

	Mean % Change from Baseline	Standard Deviation
Lumbar spine at Month 12	5.1	3.5
Total hip at Month 12	2.9	3.5
Femoral neck at Month 12	2.3	4.6
Lumbar spine at Month 24	6.3	4.9
Total hip at Month 24[a]	4.7	6.3
Femoral neck at Month 24	3.1	4.8

[a] The mean of percent change from baseline at hip trochanter is used as the assumption for total hip.

Table 2. Mean Percent Change From Baseline in BMD at Month 12 for Romosozumab

	Mean % Change from Baseline	Standard Deviation
Lumbar spine at Month 12	13.1	6.0
Total hip at Month 12	6.0	4.2
Femoral neck at Month 12	5.5	4.6

3.2.3 Simulation and Power

Simulation parameters

The assumed distributions for nonvertebral fractures and new vertebral fractures are exponential with the yearly rates specified previously. A binomial distribution with probability of 1/3 is assumed for the likelihood of a new vertebral fractures being a clinical vertebral fracture. Additionally, the censoring distribution is assumed exponential with a rate of 10% for the first year and a yearly rate of 8% thereafter. Enrollment is assumed to be completed in 34 months. As stated previously, the minimum follow-up is 24 months and all subjects will be followed until the 330th subject has a clinical fracture. A log-rank test is used for the clinical and nonvertebral fracture endpoints and a X^2 (no continuity correction) test is used for the new vertebral fracture endpoint. All tests are 2-sided and results are based on 10,000 iterations.

Power

The power for the clinical fracture and new vertebral fracture endpoint after accounting for the multiplicity adjustment are 94% and 95%. There is 91% power that both primary endpoints will be significant at the 5% level (2-sided) after accounting for multiplicity. Based on the simulation, the median time of the primary analysis in the simulation is after the last enrolled subject completes the Month 24 visit. The median number of subjects having clinical fractures in the simulation at the time of the primary analysis is 396. Note that the number of clinical fractures at the time of the primary analysis will exceed 330 when 330 subjects with confirmed clinical fractures occur before the Month 24 analysis.

If both of the primary endpoints are significant under the Hochberg procedure, each of the following secondary DXA BMD endpoints will be tested hierarchically at 0.05 according to the following sequence: percent change from baseline in BMD at lumbar spine at Month 24, percent change from baseline in BMD at total hip at Month 24, percent change from baseline in BMD at femoral neck at Month 24, percent change from baseline in BMD at lumbar spine at Month 12, percent change from baseline in BMD at total hip at Month 12, and percent change from baseline in BMD at femoral neck at Month 12. The power for each DXA BMD endpoints is $>$ 99% using the 2-sample *t*-test.

If all preceding endpoints are significant, the nonvertebral fracture endpoint will be tested using a group sequential approach at the primary analysis and the final analysis based on a 1-sided test ($\alpha = 0.025$). The Lan-DeMets alpha spending function (Lan and DeMets,1983) that approximates a Pocock boundary (Pocock, 1977), (0.025*LN(1+(EXP(1) -1)*information fraction)), will be used to determine the significance level at the time of the primary analysis. Based on the number of subjects with nonvertebral fractures at the time of the primary analysis out of the total 440 planned, which represents the information fraction in the alpha spending function, the significance level will be calculated at the time of primary analysis. For example, if the information fraction is 80% at time of the primary analysis, the significance level is 0.0216. Based on the simulation, the power for nonvertebral fracture, both primary endpoints, and DXA BMD endpoints at Month 12 and Month 24 all being significant is approximately 78% at the primary analysis.

If the significance of nonvertebral fracture is not demonstrated at the time of the primary analysis, the nonvertebral fracture endpoint will be tested again using a 1-sided test at the time of the final analysis. Based on the alpha level spent at the primary analysis, the significance level at final will be determined. For example, if the information fraction is 80% at the primary analysis (ie, the significance level is 0.0216 at primary), the significance level at the final analysis will be 0.0119 using software EAST 5.4 or R gsDesign package. Based on the simulation, the power for nonvertebral fracture being significant at the primary or final analysis, and both primary endpoints and DXA BMD endpoints at Month 12 and Month 24 being significant at primary analysis is approximately 84%.

Monitoring of Blinded Fracture Rates

The sponsor will monitor the pooled, blinded clinical fracture rate; if it is lower than expected, the sample size may be modified.

4. Study Endpoints and Covariates

4.1 Primary Endpoints

During the primary analysis period:

- Subject incidence of clinical fracture (nonvertebral fracture and clinical vertebral fracture) at primary analysis
- Subject incidence of new vertebral fracture through Month 24

4.2 Secondary Endpoints

The secondary efficacy endpoints include the following:

During the primary analysis period:

- Subject incidence of nonvertebral fracture at primary analysis
- Subject incidence of all fractures (nonvertebral fracture and new or worsening vertebral fracture) at primary analysis
- Subject incidence of new or worsening vertebral fracture through Month 24
- Subject incidence of major nonvertebral fracture (pelvis, distal femur, proximal tibia, ribs, proximal humerus, forearm, and hip) at primary analysis
- Subject incidence of hip fracture at primary analysis
- Subject incidence of multiple new or worsening vertebral fractures through Month 24
- Subject incidence of clinical fracture (nonvertebral fracture and clinical vertebral fracture) through Month 24

- Subject incidence of nonvertebral fracture through Month 24
- Subject incidence of hip fracture through Month 24
- Subject incidence of clinical vertebral fracture through Month 24
- Percent change from baseline in BMD at the lumbar spine, total hip, and femoral neck at Months 24, and 36

During the 12-month double-blind ALN-controlled study period:

- Subject incidence of clinical fracture (nonvertebral fracture and clinical vertebral fracture) through Month 12
- Subject incidence of new vertebral fracture through Month 12
- Subject incidence of all fractures (nonvertebral fracture and new or worsening vertebral fracture) through Month 12
- Subject incidence of nonvertebral fracture through Month 12
- Subject incidence of hip fracture through Month 12
- Subject incidence of major osteoporotic fracture (hip, forearm, humerus, and clinical vertebral) through Month 12
- Subject incidence of clinical vertebral fracture through Month 12
- Percent change from baseline in BMD at the lumbar spine, total hip, and femoral neck at Month 12

For the overall study period:

- Subject incidence of nonvertebral fractures at final analysis
- Subject incidence of major nonvertebral fracture (pelvis, distal femur, proximal tibia, ribs, proximal humerus, forearm, and hip) at final analysis
- Subject incidence of hip fracture at final analysis

4.3 Safety Endpoints

The safety endpoints include the following:

During the 12-month double-blind ALN-controlled study period:

- Subject incidence of adverse events by system organ class and preferred term
- Changes from baseline in laboratory assessments (serum chemistry and hematology) and the shifts from baseline to the worst value between baseline and Month 12
- Changes from baseline in vital signs
- Subject incidence of anti-romosozumab antibodies

During the primary analysis period:

- Subject incidence of adverse events by system organ class and preferred term
- Changes from baseline in laboratory assessments (serum chemistry and hematology) and the shift from baseline to the worst value between baseline and primary analysis
- Changes from baseline in vital signs
- Subject incidence of anti-romosozumab antibodies

During the overall study period:

- Subject incidence of adverse events by system organ class and preferred term
- Changes from baseline in vital signs

4.4 Exploratory Endpoints

The exploratory endpoints include the following:

During the 12-month double-blind ALN-controlled study period:

- Subject incidence of new or worsening vertebral fracture through Month 12
- Subject incidence of major nonvertebral fracture (pelvis, distal femur, proximal tibia, ribs, proximal humerus, forearm, and hip) through Month 12
- Subject incidence of multiple new or worsening vertebral fractures through Month 12
- Actual value in PRO and ClinRO measures (OPAQ SV, EQ-5D-5L, LAD, and BPI worst pain) at Months 6 and 12
- Change from baseline in PRO and ClinRO measures (OPAQ SV, EQ-5D-5L, LAD, and BPI worst pain) at Months 6 and 12
- Actual value in PRO and ClinRO measures after experiencing a nonvertebral or clinical vertebral fracture (OPAQ SV, EQ-5D-5L, LAD, and BPI worst pain) at reporting of the nonvertebral or clinical vertebral fracture and at 1, 2 and 3 months afterwards
- Change from pre-fracture baseline in PRO and ClinRO measures after experiencing a nonvertebral or clinical vertebral fracture (OPAQ SV, EQ-5D-5L, LAD, and BPI worst pain) at reporting of the nonvertebral or clinical vertebral fracture and at 1, 2 and 3 months afterwards
- Proportion of subjects with a clinically meaningful improvement in worst pain (defined as a 2-point improvement in the BPI worst pain scale compared with the fracture reporting visit) at 1, 2, and 3 months after reporting of a nonvertebral or clinical vertebral fracture
- Change from baseline in height at Month 12

During the primary analysis period:

- Subject incidence of major osteoporotic fracture (hip, forearm, humerus, and clinical vertebral) at primary analysis
- Actual value in PRO and ClinRO measures (OPAQ SV, EQ-5D-5L, LAD, and BPI worst pain) at Months 18, 24, 30, and 36
- Change from baseline in PRO and ClinRO measures (OPAQ SV, EQ-5D-5L, LAD, and BPI worst pain) at Months 18, 24, 30, and 36
- Change from baseline in height at Month 24

During the month 12 to 24 ALN study period

- Subject incidence of new vertebral fractures between Month 12 and Month 24
- Subject incidence of clinical fracture (nonvertebral fracture and clinical vertebral fracture) between Month 12 and Month 24
- Subject incidence of nonvertebral fracture between Month 12 and Month 24
- Subject incidence of hip fracture between Month 12 and Month 24
- Subject incidence of clinical vertebral fracture between Month 12 and Month 24

4.5 Imaging and PK/BTM/Biomarker Sub-study Endpoints

During the 12-month double-blind ALN-controlled study period:

- Romosozumab serum concentrations at Day 1, Months 1, 3, 6, 9, and 12
- Percent change from baseline in P1NP, BSAP, OC, and sCTX at Months 1, 3, 6, 9, and 12
- Percent change from baseline in iPTH and sclerostin at Months 1, 3, 6, 9, and 12
- For subjects participating in the imaging components:
    - Percent change from baseline in integral (total) and trabecular vBMD at the lumbar spine by QCT at Months 6, and 12
    - Percent change from baseline in lumbar spine strength as assessed by FEA at Months 6, and 12
    - Percent change from baseline in DXA BMD at the lumbar spine, total hip, and femoral neck at Month 6

During the primary analysis period:

- Percent change from baseline in P1NP and sCTX at Months 15, 18, 24, and 36
- Percent change from baseline in iPTH and sclerostin at Months 15, 18, 24, and 36
- For subjects participating in the imaging components:
    - Percent change from baseline in integral (total) and trabecular vBMD at the lumbar spine by QCT at Months 24

- Percent change from baseline in lumbar spine strength as assessed by FEA at Months 24
- Percent change from baseline in DXA BMD at the lumbar spine, total hip, and femoral neck at Month 18

4.6 Planned Covariates

All analyses assessing treatment effect of new vertebral fracture, clinical fracture or nonvertebral fracture will include the stratification factor (age < 75 or ≥ 75 years), presence or absence of severe vertebral fracture at baseline, and baseline total hip BMD T-score as main covariates in the model. Additional covariates will be analyzed separately and simultaneously, if appropriate, as exploratory analyses. The additional covariates of interest include the following:

- baseline body mass index (BMI)
- years since menopause at baseline
- prior (history) nonvertebral osteoporotic fracture (yes, no)
- geographic region:
    - Western Europe and New Zealand/Australia
    - Central and Eastern Europe and Middle East
    - Asia Pacific and South Africa
    - North America (including Canada, and United States)
    - Central/Latin America
- race/ethnicity (white and non-white)
- prior use of bone-specific therapeutic agents (yes, no)
- 10-year probability of major osteoporotic fracture with BMD (FRAX, as calculated by third-party vendor)
- baseline lumbar spine BMD T-score

5. Hypotheses and/or Estimation

The primary clinical hypotheses are that romosozumab treatment for 12 months followed by ALN treatment compared with ALN treatment alone is effective in reducing the incidence of clinical fracture (nonvertebral fracture and clinical vertebral fracture) and new vertebral fracture in postmenopausal women with osteoporosis. It is expected that romosozumab treatment for 12 months followed by ALN treatment will reduce the incidence of clinical fractures by 30% and the incidence of new vertebral fractures by 50% compared with the control group (ALN alone).

The primary safety hypothesis is that romosozumab treatment for 12 months is well tolerated in postmenopausal women with osteoporosis.

6. Definitions

6.1 Basic Definitions

Investigational Product (IP)

Romosozumab, ALN and the respective matched placebo

Interactive Voice Response System (IVRS)

The system used to assign screened subjects to randomized treatment as well as to manage the supply of double-blind romosozumab and ALN, as well as the open-label ALN to the site and track subject study termination data

6.1.1 Study Points of Reference

Baseline

Baseline is the closest recorded measurement before the administration of the first dose of investigational product. If the measurement is done on the same day as the first dose and the exact measurement time relative to the first dose is unknown, it will be assumed the measurement is done before the administration of the first dose of investigational product. If a subject does not receive investigational product, baseline is the closest recorded measurement on or before the enrollment (randomization) date.

Note: If baseline result from lateral spine x-ray, DXA, or QCT assessment is not available, the result assessed on or before Study Day 14 will be considered baseline.

Pre-fracture Baseline for PRO/ClinRO Endpoints

Pre-fracture baseline for PRO/ClinRO endpoints is the closest recorded PRO/ClinRO measurements before the visit of reporting a nonvertebral fracture or clinical vertebral fracture.

PRO/ClinRO Endpoints at Report of Fracture

PRO/ClinRO Endpoints at Report of Fracture is the first recorded PRO/ClinRO measurements within 45 days after the date of the reported non-vertebral fracture or clinical vertebral fracture.

Study Day 1

The first day of investigational product administration or the day of randomization for subjects who were not administered any dose of investigational product

Study Day

The number of days from Study Day 1, inclusive:

Study Day = (Date of Interest – Date of Study Day 1) + 1

Visit Windows

Based on protocol, all monthly study visits through Month 12, during the 12-month double-blind ALN-controlled study period, have a ± 7-day window. The study visits during the open-label ALN study period , have a ± 14-day window. Study procedures for a specific visit may be completed on multiple days as long as all the procedures are completed within the visit window. To allow for variations in scheduling, the analysis visit windows defined in section 13.1.1 will be used to assign evaluations to a most appropriate nominal visit for analysis and summarization.

6.1.2 Study Dates

Informed Consent Date

The date on which a subject signed the informed consent

Enrollment (Randomization) Date

The date on which a subject is assigned to **one** of the treatments through the IVRS

End of Primary Analysis Period Date

The date on which clinical fracture events have been confirmed for at least 330 subjects and after all randomized subjects have had the opportunity to complete the Month 24 visit. For subjects who withdraw before this date, then their End of Study Date is used.

End of Final Analysis Period Date

If the study will continue to the final analysis, the date on which nonvertebral fracture events have been confirmed for at least 440 subjects if statistical significance of non-vertebral fractures is not demonstrated at primary analysis. For subjects who withdraw before this date, then their End of Study Date is used.

End of Study Date

The date recorded on the End of Study eCRF

First Dose Date

The date of administration of first dose of investigational product (ie, the first date recorded on the Investigational Product Administration eCRF for romosozumab or placebo with volume > 0 or the first dose date of ALN or placebo in

Alendronate/Placebo Dispensation eCRF with the number of tablets taken > 0, whichever is earlier), which may or may not be the same as the randomization date

Last Double-blind SC Dose Date

The date of administration of the last SC investigational product (ie, the last date recorded on the Investigational Product Administration eCRF for romosozumab or placebo with volume > 0)

Last Double-blind Oral Dose Date

The date collected on the End of Blinded Alendronate Administration eCRF

Last Dose Date in Double-blind Period

The date of last double-blind SC dose date or the last double-blind oral dose date, whichever is later

First Open-label ALN Dose Date

The date collected on the First Dose of open-label Alendronate Dispensation eCRF

First Open-label ALN Dispensation Date

The date of dispensation of ALN collected on the Month 12 Alendronate Dispensation eCRF

Last Open-label ALN Dose Date

The date of last alendronate administration recorded in End of Alendronate Administration eCRF

End of Double-blind Period Date

End of the double-blind period date is defined as the following:

1. First open-label ALN dose date if this date is prior to or on Study Day 366 + 30 days
2. Month 12 visit date if subject did not receive the open-label ALN prior to or on Study Day 396 and this date is prior to or on Study Day 396 days
3. End of study day or Study Day 396, whichever occurs first, for subjects who did not complete the Month 12 visit prior to study day 396 days

The Month 12 visit date is defined as the latest assessment collected at Month 12.

Start of Open-label Period Date

For subjects entering in the ALN open-label period, the end of double-blind period date is the beginning of the open-label period date. All scheduled

assessments occurring on this date are attributed to the double-blind period. The open-label ALN dose on the day is considered as occurring in the open-label follow-up period. Any adverse event or concomitant medication with a start date on this date is considered having started in the open-label follow-up period.

End of 24-Month ALN Study Period Date

End of the 24-Month period date is defined as the following:

1. Month 24 visit date if this date is prior to or on Study Day 731 + 30 days
2. End of study date or Study Day 761, whichever occurs first, for subjects who did not complete the Month 24 visit prior to Study Day 761 days

The Month 24 visit date is defined as the latest assessment collected at Month 24.

Last Dose Date

The last dose date in double-blind or open-label period, whichever is later

Wrong Dose Date in Double-blind Period

For subjects who are randomized to ALN and accidentally receive romosozumab, the wrong dose date is when they receive the 1st SC administration of romosozumab in the double-blinded period. For subjects who are randomized to romosozumab and receive oral ALN in the double-blinded period, if the wrong oral IP box is dispensed on Day 1, the wrong dose date is the 1st oral dose date; otherwise, the wrong dose date is when the 1st wrong oral IP box is dispensed because the oral ALN dosing dates are not collected in eCRF.

6.1.3 Study Time Intervals

On-Study Period

The time period from the enrollment date to the end of study date, inclusive

Double-blind Period

The time period from the enrollment date to the end of double-blind period date, inclusive

Open-label ALN Period

The time period from the end of double-blind period date to the end of study date, inclusive

Month 12 to Month 24 ALN Study Period

For subjects entering in the ALN open-label period, the time period from the end of double-blind period date to the Month 24 visit date, study day 761, or end of study date, whichever is the earliest, inclusive.

The Month 24 visit date is defined as the latest assessment collected at Month 24.

24-month ALN Study Period

The time period from the enrollment date to the Month 24 visit date, study day 761, or end of study date, whichever is the earliest, inclusive

Primary Analysis Period

The time period from the enrollment date to the end of primary analysis period date, inclusive

Overall Study Period

If the study will continue to the final analysis, the overall study period is defined as the time period from the enrollment date to the end of final analysis period date, inclusive.

6.1.4 Fracture Risk Assessment Tool: 10-year Probability of Major Osteoporotic Fracture (FRAX)

FRAX is a computer-based algorithm (http://www.shef.ac.uk/FRAX) that provides country/ethnicity-specific models for the assessment of fracture probability in men and women [Kanis et al, 2001,2008]. The approach uses clinical risk factors to estimate 10-year probability of a major osteoporotic fracture (hip, clinical spine, forearm or humerus) or of a hip fracture alone. The estimate of probability can be calculated with clinical risk factors alone, or additionally with baseline femoral neck BMD. The clinical risk factors used for the calculation include sex, age, body mass index (BMI), a prior fragility fracture, parental history of hip fracture, current tobacco smoking, ever long-term use of oral glucocorticoids, rheumatoid arthritis, other causes of secondary osteoporosis, and daily alcohol consumption of 3 or more units daily.

6.2 Subject Disposition

Enrolled (Randomized)

Individuals are considered enrolled if they have been assigned a randomization number. Enrolled individuals are referred to as “subjects”.

Exposed to Investigational Product

Subjects are defined as exposed if they have a value for the sum of investigational product dose that exceeds zero.

Enrolled Into Sub-study

Randomized subjects who sign the inform**ed** consent for the sub-study are considered enrolled into the sub-study.

6.3 Arithmetic Calculations

Age at Randomization Date

Number of whole years from a subject's birth date to the randomization date as recorded on the eCRF.

Estimated Glomerular Filtration Rate (eGFR) (mL/min) Using Modification of Diet in Renal Disease (MDRD) for Females

Conventional units (creatinine as mg/dL; age in years):
eGFR (mL/min/1.73 m2) = 175 × (Serum creatinine)-1.154 x [Age]-0.203 × (0.742 if female) × (1.212 if African American)

SI units (creatinine as μmol/L; age in years):
eGFR (mL/min/1.73 m2) = 175 × (Serum creatinine/88.4)-1.154 x [Age]-0.203 × (0.742 if female) × (1.212 if African American)

Percent Change From Baseline

The change from baseline divided by baseline value and multiplied by 100:

(Change From Baseline / Baseline) * 100

Subject Incidence Rate for Adverse Events

The subject incidence rate for a given event in a given time period (double-blind period, primary analysis period, or overall study period) is defined as the number of subjects with ≥ 1 reported occurrence of the event divided by the number of subjects who are at risk for having the event in the beginning of the given time period. For subjects with multiple occurrences of the same event in a given period, the event will only be counted once per subject.

Subject Years On-Study in Double-blind Period

Defined for a given subject as the number of days between the first dose date to end of double-blind period date, inclusive, divided by 365.25 (ie, [End of Double-blind Period Date – First Dose Date + 1] / 365.25). The subject years on study will be summed over all subjects within a treatment group. For subjects who never received a dose of investigational product, the randomization date will be used.

Subject Years On-Study in Primary Analysis Period

Defined for a given subject as the number of days between the first dose date to end of primary analysis period date, inclusive, divided by 365.25 (ie, [End of Primary Analysis Period Date – First Dose Date +1] / 365.25). The subject years on study will be summed over all subjects within a treatment group. For subjects who never received a dose of investigational product, the randomization date will be used.

Subject Years On-Study in Overall Study Period

Defined for a given subject as the number of days between the first dose date to end-of-study date, inclusive, divided by 365.25 (ie, [End of Study Date - First Dose Date +1] / 365.25). The subject years on study will be summed over all subjects within a treatment group. For subjects who never received a dose of investigational product, the randomization date will be used.

Time to First Event in Double-blind Period

Time to event is calculated as the elapsed time interval in days between the occurrence date of the event of interest or censorship and the reference date:

Time Interval = Date of the Event or Censorship - Reference Date +1

Note for fracture endpoints, the date of event is the date that confirms the fracture provided by the central imaging reader (eg, date of x-ray used by the central imaging vendor to verify the fracture, or surgical report or discharge summary if x-ray date is not available). For adverse events, nonvertebral fractures, and other events, which can be assessed at any time, subjects who did not have any event will be censored on the date of last evaluation for the event, which is the End of Double-blind Study Period; for vertebral fractures, which have to be identified on scheduled assessments, subjects who did not have any event will be considered censored at the last post-baseline assessment or **d**ay 1 if there were no post-baseline assessment.

The reference date will be the randomization date for efficacy events or first dose date for safety endpoints.

Time to First Event in 24-month Study Period

The time to event in the 24-month study period is defined similarly as for time to first event in double-blind period. Subjects who did not have any event will be censored on the date of last evaluation for the event, which is the end of 24-month ALN study period date.

Time to First Event in Month 12 to Month 24 Study Period

Time to first event in Month 12 to Month 24 study period is defined similarly as for time to first event in 24-month study period. The reference date will be the start of open-label period.

Time to First Event in Primary Analysis Period

The time to event in the primary analysis study period is defined similarly as for time to first event in double-blind period. Subjects who did not have any event will be censored on the date of last evaluation for the event, which is the end of primary analysis period date.

Time to First Event in Overall Study Period

The time to event in the overall study period is defined similarly as for time to first event in double-blind period. Subjects who did not have any event will be censored on the date of last evaluation for the event, which is the End of Study Date recorded on the End of Study eCRF.

6.4 Fracture-related Study Endpoints

All the fracture related efficacy endpoints are based on the results from the central imaging vendor analysis transferred to the Amgen database.

6.4.1 Trauma Severity Definitions

Low Trauma Severity

Assessed by the investigator and collected on the Clinical Non-vertebral Fracture Summary eCRF for each clinical fracture event and includes

- Fall from standing height or less than 20 inches
- Minimal or moderate trauma other than a fall
- Unknown/don’t know

High Trauma Severity

Assessed by the investigator and collected on the Clinical Non-vertebral Fracture Summary eCRF for each clinical fracture event indicating severe trauma other than a fall

Pathologic Fracture

Assessed by the investigator and collected on the Clinical Non-vertebral Fracture Summary eCRF for each clinical fracture event occurring from a pathology other than osteoporosis is deemed a pathologic fracture

6.4.2 Vertebral Fracture

The types of vertebral fractures below are defined based on an assessment of spinal radiographs using Genant Semiquantitative Scoring Method (Genant et al, 1993), which is described in detail in Section 13.2.

Imputation of Genant Grade for a Vertebra

Because in an adult, a subject's vertebral fracture can only get worse or at best remain at the same severity over time, for the evaluation of prevalent vertebral fracture, any vertebra(e) with a missing grade at baseline will be assumed to have had a grade of 0 if the subsequent x-ray shows a grade of 0 for the same vertebra(e).

Prevalent Vertebral Fracture Status at Baseline

Most subjects enrolled to this study will have at least one vertebral fracture. The most severe Genant semiquantitative grade from T4 to L4 at baseline is used to define the vertebral fracture status at baseline. A subject has a prevalent vertebral fracture if any vertebra from T4 to L4 has a grade of ≥ 1 at baseline. A subject does not have a prevalent vertebral fracture when all 13 grades from T4 to L4 are 0 on the first evaluable spinal radiograph collected during the study. Otherwise, the subject will have an unknown status for prevalent vertebral fracture.

Presence of Severe Vertebral Fracture at Baseline

A subject has a prevalent vertebral fracture if any vertebra from T4 to L4 has a grade of 3 at baseline. Otherwise, a subject is considered absence of severe vertebral fracture at baseline.

New Vertebral Fracture

A new vertebral fracture is identified when there is ≥ 1 grade increase from the previous grade of 0 in any vertebra from T4 to L4.

Worsening Vertebral Fracture

A worsening vertebral fracture is identified when there is ≥ 1 grade increase from the previous grade of ≥ 1 in any vertebra from T4 to L4.

New or Worsening Vertebral Fracture

A new or worsening vertebral fracture is identified when there is ≥ 1 grade increase from the previous grade in any vertebra from T4 to L4.

Clinical Vertebral Fracture

A clinical vertebral fracture is a new or worsening vertebral fracture assessed at either a scheduled or unscheduled visit and associated with any signs and/or symptoms of back pain indicative of a fracture, regardless of trauma severity or whether it is pathologic.

Signs and symptoms indicative of a vertebral fracture will be assessed using the Clinical Vertebral Fracture Back Pain eCRF. Data from this eCRF will be combined with the post-baseline lateral spine x-ray data from the central imaging

vendor by taking all back pain assessments prior to or on the date of the x-ray and the first eCRF record on or after the date of the x-ray and prior or on to the next x-ray date. If the x-ray date corresponds exactly with an eCRF record that indicates no back pain was present at the time of the x-ray, the subsequent back pain assessment will also be considered, so long as it is on or before the next x-ray. If any selected records indicate the subject experienced back pain consistent with a vertebral fracture, then the vertebral fracture will be considered as associated with signs and/or symptoms indicative of a vertebral fracture, and therefore a clinical vertebral fracture. If all selected records indicate that the subject did not experience back pain deemed consistent with a vertebral fracture by the investigator around the time of the fracture, then the fracture will not be considered a clinical vertebral fracture.

Multiple New or Worsening Vertebral Fractures

A subject has multiple new or worsening vertebral fractures when there are ≥ 2 vertebrae from T4 to L4 with ≥ 1 grade increase from the previous grade. The multiple new or worsening vertebral fractures need not occur at the same visit.

6.4.3 Nonvertebral Fracture (Osteoporotic)

Nonvertebral Fracture (Osteoporotic)

Defined as a fracture present on a copy of radiographs or other diagnostic images such as computerized tomography (CT) or magnetic resonance imaging (MRI) confirming the fracture within 14 days of reported fracture image date on the eCRF (if day of the month is unknown, but month and year is known, then fracture images within the same month can be used to confirm), and/or documented in a copy of the radiology report, surgical report, or discharge summary, excluding skull, facial, mandible, cervical vertebrae, thoracic vertebrae, lumbar vertebrae, metacarpus, finger phalanges, and toe phalanges. In addition, fractures associated with high trauma severity or pathologic fractures will be excluded.

Hip Fracture

A subset of nonvertebral fractures including femoral neck, femur intertrochanter, femur subtrochanter, and “Other” location that are further specified as hip, pertrochanteric, or pertrochanteric femur.

Distal Femur Fracture (ie, Femur Excluding Hip)

A subset of nonvertebral fractures including femur distal and femur midshaft

Wrist Fracture

A subset of nonvertebral fractures including radius distal and ulna distal

Forearm Fracture

A subset of nonvertebral fractures including radius, radius proximal, radius shaft, radius distal, ulna, ulna proximal, ulna shaft, and ulna distal

Rib Fracture

A subset of nonvertebral fractures including ribs

Humerus Fracture

A subset of nonvertebral fractures including humerus proximal, humerus shaft, and humerus distal

Proximal Humerus Fracture (ie, Humerus Excluding Elbow)

A subset of nonvertebral fractures including humerus proximal and humerus shaft

Pelvic Fracture

A subset of nonvertebral fractures including sacrum, acetabulum, ilium, ischium, and pubis

Leg Fracture

A subset of nonvertebral fractures including femur midshaft, femur distal, patella, fibula, fibula proximal, fibula shaft, fibula distal, tibia, tibia proximal, tibia shaft, and tibia distal

Lower Leg With Ankle

A subset of nonvertebral fractures including fibula, fibula proximal, fibula shaft, fibula distal, tibia, tibia proximal, tibia shaft, and tibia distal

Ankle Fracture

A subset of nonvertebral fractures including fibula distal and tibia distal

Proximal Tibia Fracture (ie, Tibia Excluding Ankle)

A subset of nonvertebral fractures including tibia proximal and tibia shaft

Foot Fracture

A subset of nonvertebral fractures including metatarsus and tarsus

Major Nonvertebral Fracture

A subset of nonvertebral fractures including pelvis, distal femur (ie, femur excluding hip), proximal tibia (ie, tibia excluding ankle), ribs, proximal humerus (ie, humerus excluding elbow), forearm, and hip

Any Nonvertebral Fracture

Defined as any fractures recorded on the Clinical Non-vertebral Fracture Summary eCRF.

6.4.4 Clinical Fracture

Clinical fractures include clinical vertebral as defined in Section 6.4.2 and nonvertebral fractures (excluding skull, facial, mandible, cervical vertebrae, thoracic vertebrae, lumbar vertebrae, metacarpus, finger phalanges, and toe phalanges) as defined in Section 6.4.3 that are associated with signs and/or symptoms indicative of a fracture. Nonvertebral fractures associated with high trauma severity or pathologic fractures will be excluded.

6.4.5 All Fracture (Osteoporotic)

All osteoporotic fractures include any osteoporotic nonvertebral fractures that are not associated with high trauma severity or pathologic fractures and new or worsening vertebral fractures regardless of trauma severity or pathologic fractures.

6.4.6 Major Osteoporotic Fracture

Major osteoporotic fractures include hip, forearm, or humerus fractures that are not associated with a pathologic fracture regardless of trauma severity, and clinical vertebral fractures.

6.4.7 Historical Fracture

Any Historical Fracture

Any nonvertebral fractures recorded on the Subject Fracture History eCRF regardless of trauma severity or vertebral fracture based on baseline spinal radiograph.

Historical Osteoporotic Fracture

Any nonvertebral fracture recorded on the Subject Fracture History eCRF not including skull, facial bones, fingers, toes, spine and tailbone and not associated with known high trauma severity or pathologic fractures, or vertebral fracture based on baseline spinal radiograph

Historical Nonvertebral Fracture

Fractures recorded on the Subject Fracture History eCRF not including skull, facial bones, fingers, toes, spine, and tailbone and not associated with known high trauma severity or pathologic fractures

Historical Nonvertebral Fracture At or After Age 55

Historical nonvertebral fractures occurring at or after age 55

Historical Major Nonvertebral Fracture

A subset of historical nonvertebral fractures including the following locations: pelvis (not hip), hip, lower leg (not knee or ankle), ribs, shoulder, forearm, and wrist and not associated with known high trauma severity or pathologic fractures

Historical Fragility Fracture

Historical fragility fractures include moderate or severe vertebral prevalent fractures and nonvertebral fractures occurring at or after age 55.

6.5 PRO Measurements

OPAQ-SV Physical Function Score

The physical function score is calculated by averaging nonmissing rating for 19 standardized items (question item numbers of 1 to 19; ≥ 10 items as nonmissing; otherwise set score to missing). The OPAQ-SV physical function score ranges from 0 to 100. A higher score represents better physical function. See Section 13.3.1 for detailed scoring algorithm.

OPAQ-SV Emotional Status Score

The emotional status score is calculated by averaging nonmissing rating of 11 standardized items (question item numbers of 20 to 24, and 29 to 34; ≥ 6 items as nonmissing; otherwise set score to missing). The OPAQ-SV emotional status score ranges from 0 to 100. A higher score represents better emotional status. See Section 13.3.1 for detailed scoring algorithm.

OPAQ-SV Back Pain Score

The score is calculated by averaging nonmissing rating of 4 standardized items (question item numbers of 25 to 28; ≥ 2 items as nonmissing; otherwise set score to missing). The maximum OPAQ-SV back pain score is 100. A higher score indicates less back pain. See Section 13.3.1 for detailed scoring algorithm.

EQ-5D-5L Health State Index Score

The EQ-5D-5L health state index score is calculated using the dimension scores from Question 35 a-e. The score ranges between –0.594 and 1.0, where

1 indicates full health; 0 indicates death. A score < 0 indicates a state worse than death. See Section 13.3.2 for detailed scoring algorithm.

EQ-5D-5L VAS Score

The EQ-5D-5L VAS score ranges between 0 and 100 based on Question 36 (See Section 13.3.2 for detailed question description). The score of 0 indicates "worst imaginable health state" and 100 for "best imaginable health state."

7. Analysis Subsets

7.1 Primary Efficacy Analysis Set For Vertebral Fractures

This analysis set includes all randomized subjects who have a baseline and ≥ 1 post-baseline evaluation of vertebral fracture at or before the time point under consideration. This analysis set will additionally include subjects who have vertebrae with missing Genant semiquantitative scores at baseline and whose first post-baseline spinal radiograph shows no fracture on the same vertebrae because it can be inferred that the baseline scores would have also shown no fracture had they been available. Subjects in this subset will be analyzed according to their randomized treatment assignment, regardless of treatment received. This analysis set will be used as the primary analysis set for the following endpoints: new, new or worsening, and multiple new or worsening vertebral fractures.

7.2 Full Analysis Set

This set includes all randomized subjects. Subjects in the full analysis set will be analyzed according to their randomized treatment assignment, regardless of treatment received. The full analysis set will be used as the primary analysis set for the following endpoints: nonvertebral fracture, clinical fracture, clinical vertebral fracture, all fracture, major nonvertebral fracture, major osteoporotic fracture, and hip fracture.

7.3 Per Protocol Analysis Set for 12-month of Double-blind Period

The per protocol analysis set will only be used to analyze the following endpoints: subject incidence of clinical fracture, new vertebral fracture, and nonvertebral fracture through Month 12 as sensitivity analyses. This subset includes subjects defined in Section 7.1 (for new vertebral fractures) or Section 7.2 (for clinical fracture and nonvertebral fracture) who receive active investigational product and did not violate any of the important inclusion/exclusion criteria for subject eligibility at enrollment per study important protocol deviation documents.

During the double-blind study period, for subjects who (a) received the investigational product not matching the subject's randomized treatment group, (b) had an investigational product related important protocol violation for romosozumab /placebo or ALN/placebo during the double-blind period, or (c) received proscribed therapy on study, all vertebral fracture data collected after the first occurrence of either of the above violations will be excluded from analysis; and nonvertebral and clinical fracture endpoints analyzed as time-to-event will be censored at the time of the first violation for subjects who have not had the fracture prior to the violation date. If subjects had an investigational product related important protocol deviation during the double-blind period, the violation date is determined as follows:

- if the subject is randomized to romosozumab, the time of violation for subjects missing > 3 doses of romosozumab is defined as the planned dosing date for the 4th (missed) dose or the subject's end of study date, whichever occurs first;

- if the subject is randomized to ALN, the first deviation date for "Less than 75% or more than 125% of planned ALN/placebo doses" IPD recorded in eClinical or the subject's end of study date, whichever occurs first.

The use of proscribed treatments will result in excluding or censoring the data from the per protocol analysis set. The list of proscribed treatments is provided in protocol Section 6.5.

7.4 Per Protocol Analysis Set for 24-Month Study Period, Primary Analysis Period, and Overall Study Period

The per protocol analysis set will be used to analyze the clinical fracture, new vertebral fracture, and nonvertebral fracture through Month 24, clinical fracture and nonvertebral fracture at primary analysis, and nonvertebral fracture at final analysis (if the study continue to final analysis) as a sensitivity analysis. This subset includes subjects defined in Section 7.1 (for vertebral fracture related endpoints) or Section 7.2 (for nonvertebral related endpoints) who did not violate any of the important inclusion/exclusion criteria for subject eligibility at enrollment per study important protocol deviation documents.

For subjects who received (a) the investigational product not matching the subject's randomized treatment group during the double-blind period, (b) investigational product related important protocol violation for romosozumab/placebo or ALN/placebo during the

double-blind period, (c) investigational product related important protocol violation for open-label ALN during the open-label period of primary analysis period, or (d) proscribed therapy on study during the overall study period, all vertebral fracture data collected after the first occurrence of either of the above violations will be excluded from analysis; and nonvertebral and clinical fracture endpoints analyzed as time-to-event will be censored at the time of the first violation for subjects who have not had the fracture prior to the violation time. The use of proscribed therapies and investigational product related important protocol deviations during the double-blind period that will result in excluding or censoring the data from the per protocol analysis set is specified in Section 7.3. In addition, if subjects had an investigational product related important protocol deviation during the open-label period, the violation date is determined as the first deviation date for "Less than 75% or more than 125% of planned ALN doses" IPD recorded in eClinical or the subject's end of study date, whichever occurs first during the open-label period.

7.5 Safety Analysis Set

This safety analysis set includes all randomized subjects who receive ≥ 1 active dose of investigational product in the 12-month double-blind ALN-controlled study period. This analysis set will be used to analyze safety data for the double-blind study period, primary analysis period, and overall study period. AEs that begin on the end of double-blind period date will not be included in the 12-month double-blind study period. AEs that begin on the date of the end of primary analysis study period will be included in the primary analysis study period. These subjects will be analyzed according to their actual treatment received, where subjects who received ≥ 1 dose of romosozumab will be analyzed in the romosozumab treatment group regardless of the randomized treatment. The subject incidence rates for the primary analysis study period include all events that occurred in the double-blind study period and all events occurred on and before the end of primary analysis date for those subjects who received at least one dose of open-label ALN.

If the study continues to the finaly analysis, the subject incidence rates for the overall study period include all events that occurred in the double-blind study period and all events occurred in the open-label ALN period for those subjects who received at least one dose of open-label ALN.

7.6 Primary Efficacy Analysis Set for BMD, Height, and PRO/ClinRO Endpoints for the Double-blind ALN-controlled Study Period and Primary Analysis Period

This analysis set includes all randomized subjects who have a baseline and ≥ 1 post-baseline evaluation at or before the time point under consideration. Note that this subset could potentially be different from endpoint to endpoint due to missing data. Subjects in this subset will be analyzed according to their randomized treatment assignment, regardless of treatment received.

7.7 Analysis Set for Post Fracture PRO/ClinRO Endpoints

This analysis set includes all randomized subjects who experienced a nonvertebral or clinical vertebral fracture in the double-blind period and have a pre-fracture baseline and ≥ 1 post fracture evaluation at or before the time point under consideration. Note that this subset could potentially be different from endpoint to endpoint due to missing data. Subjects in this subset will be analyzed according to the randomized treatment assignment.

7.8 Analysis Subsets for Imaging and Pharmacokinetics (PK) / Bone Turnover Marker (BTM) / Biomarker Sub-study

7.8.1 Sub-study Imaging Efficacy Analysis Set

The respective subset includes all randomized subjects who enrolled in the sub-study and have a baseline and ≥ 1 post-baseline evaluation at or before the time point under consideration and will be used to evaluate vBMD, bone strength measure by FEA and DXA BMD. Subjects will be analyzed according to their randomized treatment assignment.

7.8.2 Sub-study BTM Efficacy Analysis Set

The subset includes all randomized subjects who enrolled in the sub-study who have a baseline and ≥ 1 post-baseline reported BTM result at or before the time point under consideration and will be used to evaluate P1NP, BSAP, OC and sCTX. Subjects will be analyzed according to their randomized treatment assignment, regardless of treatment received.

7.8.3 Sub-study BTM Safety Analysis Set

The subset includes all randomized subjects who enrolled in the sub-study who receive ≥ 1 dose of investigational product, have a baseline and ≥ 1 post-baseline reported BTM result at or before the time point under consideration and will be used to evaluate iPTH and serum sclerostin level. Subjects will be analyzed according to the actual reatment received, where subjects who received ≥ 1 dose of romosozumab will be analyzed in the romosozumab treatment group regardless of the randomized treatment.

7.8.4 Sub-study PK Analysis Set

The subset includes all randomized subjects who enrolled in the PK sub-study, receive ≥ 1 dose of romosozumab, and have ≥ 1 reported result. Subjects will be analyzed according to their actual treatment received.

7.9 Subgroup Analyses

The new vertebral fracture through Month 12 or Month 24, clinical fracture and nonvertebral fracture through Month 12 or at primary analysis, will be analyzed within each of the following subgroups:

- Age (< 75 years, ≥ 75 years)
- Presence or absence of severe vertebral fracture at baseline (based on most severe Genant semiquantitative grade at baseline on screening spinal radiograph)
- Number of prevalent vertebral fractures at baseline (0-1, 2, ≥ 3; based on screening spinal radiograph)
- Race (White and non-White)
- Geographic region
  - Western Europe and New Zealand/Australia
  - Central and Eastern Europe and Middle East
  - Asia Pacific and South Africa
  - North America
  - Central/Latin America
- Central/Latin America and all regions excluding Central/Latin America
- Baseline lumbar spine BMD T-score (≤ -3, > -3 and ≤ -2.5, > -2.5)
- Baseline total hip or femoral neck BMD T-score ≤ -3 vs both total hip and femoral BMD T-score > -3

- Baseline BMI (tertiles)
- FRAX score for major osteoporotic fracture (tertiles)
- History of nonvertebral fracture at age ≥ 55 years (yes, no)

The percent changes from baseline in lumbar spine BMD and total hip BMD at month 12 and month 24 will be analyzed by the following subgroups:

- Age (< 75 years, ≥75 years)
- Geographic region
- Baseline lumbar spine BMD T-score (≤ -3, > -3 and ≤ -2.5, > -2.5; for lumbar spine BMD analysis only)
- Baseline total hip BMD T-score (≤ -3, > -3 and ≤ -2.5, > -2.5; for total hip BMD analysis only)

These subgroups, except for age strata, will be re-examined for appropriateness and may be re-categorized or omitted (due to small sample size, for example, if there are < 10% of subjects within a subgroup) before unblinding. The analyses of these subgroups will be exploratory in nature.

7.10 Efficacy Analysis Set in Month 12 to 24 Study Period

This set includes all randomized subjects who received at least one open-label ALN dose. Subjects in the analysis set will be analyzed according to their randomized treatment assignment, regardless of treatment received. The analysis set will be used as the primary analysis set for the following endpoints between Month 12 and Month 24: nonvertebral fracture, clinical fracture, clinical vertebral fracture, and hip fracture.

7.11 Efficacy Analysis Set for Vertebral Fracture in Month 12 to 24 Study Period

This analysis set includes all randomized subjects who received at least one open-label ALN dose and have evaluation of vertebral fracture at Month 12 and at or before Month 24. This analysis set will additionally include subjects who have vertebrae with missing Genant semiquantitative scores prior to Month 12 and whose first spinal radiograph after Month 12 shows no fracture on the same vertebrae because it can be inferred that the baseline and Month 12 scores would have also shown no fracture had they been available. This analysis set will be used as the primary analysis set for new vertebral fracture between Month 12 and Month 24.

8. Interim Analysis and Early Stopping Guidelines

No formal interim analysis is planned for this study prior to the evaluation of the primary endpoints.

An external, independent DMC will monitor unblinded safety data on an ongoing basis throughout the entire study period and, if a potential safety signal is identified, may consider efficacy data in order to assess the risk/benefit profile of romosozumab. DMC members will have access to treatment assignments if knowledge of treatment assignment at the individual level is essential to evaluate safety. To minimize the potential introduction of bias, these individuals will not have direct contact with the study site personnel or subjects. An independent statistical service provider will generate unblinded reports for review by the DMC. If at any time there are safety concerns, the DMC will communicate the concerns to a representative from Amgen senior management. The DMC will convene approximately every 3 to 6 months. The start date will depend on the subject accrual rates. Roles and groups requiring access to restricted and interim data at the time of the primary analyses and any time before initial database lock will be pre-specified to ensure the integrity of the trial.

Records of all meetings will be maintained by the DMC for the duration of the study. Records of all meetings will be stored in the Amgen official document management system Trial Master File at the conclusion of the study. Further details are provided in the DMC Charter.

Staff from the Pharmacokinetics and Drug Metabolism group will be unblinded to treatment assignments before study final unblinding to perform exposure-response analyses. These individuals will ensure potentially unblinding data are not distributed to blinded individuals until the study is unblinded.

Early access to data for exposure-response analysis is solely for the purpose of preparing datasets and statistical programs (ie, administrative processing). Any in-process results from this early access will not be distributed and do not constitute an interim analysis.

9. Data Screening and Acceptance

9.1 Data Handling and Electronic Transfer of Data

All data for this study will be received from Amgen's data management department and will be housed within the Electronic Data Capture (EDC) database, RAVE. All screening and on-study blood samples will be processed by the central laboratory (with the exception of the serum electrophoresis samples that are processed/analyzed by a local laboratory) and will be electronically transferred to the Amgen database. All imaging data (eg, X-ray, QCT, FEA, and DXA), will be submitted to the central imaging vendor for final analysis. The results from the central imaging vendor analysis will be electronically transferred to the Amgen database. All other data will be captured on the eCRF.

An Analysis Dataset for Pharmacokinetics Concentrations (ADPC) will be provided to the Clinical Pharmacology, and Modeling and Simulation group from Biostatistics.

9.2 Handling of Missing and Incomplete Data

Subjects may have missing specific data points for a variety of causes. In general, data may be missing due to a subject's early withdrawal from study, a missed visit, or non-evaluability of a specific clinical measurement at its planned clinical visit. Unless specified, no imputation will be used. The general procedures outlined below describe what will be done when a data point is missing.

9.2.1 Genant Semiquantitative Grades

Missing semiquantitative grades will not be imputed to determine vertebral fracture status.

9.2.2 Vertebral Fracture Assessment

Because in an adult, a vertebral fracture can only get worse or at best remain at the same severity over time, the Genant semiquantitative grade for a vertebra can only increase or remain the same. In other words, once a vertebral fracture is identified, the subsequent spinal radiographs will always show fracture. Therefore, any missing post-baseline vertebral fracture status due to missing spinal x-ray assessment will be imputed using the status from the last nonmissing post-baseline visit (ie, last observation carried forward [LOCF]).

9.2.3 DXA or QCT Endpoints

Missing baseline values for endpoints by DXA or QCT at any anatomical site will not be imputed. Missing post-baseline values will be imputed using LOCF approach (by carrying forward the last nonmissing post-baseline value prior to the missing value from the same anatomical site). For the open-label period endpoints, only the open-label period value will be used for imputation, the values in the double-blind period will not be carried forward into open-label period. The LOCF imputation will be done up to Month 36. Missing post-baseline data will not be imputed when repeated measures model is employed as a sensitivity analysis.

If a subject has values from different DXA machine types (ie, Hologic and Lunar) or QCT machine types only those values that are collected from the same machine type as the baseline will be used for analyses and imputation. For anatomical sites that can be measured on different body sides (ie, left and right), only those values that are collected from the same body side as the baseline will be used for analyses and imputation.

9.2.4 Bone Turnover Markers/Biomarkers and PK

Missing bone turnover maker or biomarkers (either baseline or post-baseline values) or PK will not be imputed. Any values below the lower limit of quantification will be imputed using the lower limit of quantification for analysis.

9.2.5 PRO Measurements

9.2.5.1 OPAQ-SV

Missing baseline or post-baseline OPAQ scores will not be imputed.

9.2.5.2 EQ-5D-5L

Missing baseline or post-baseline EQ-5D-5L scores will not be imputed.

9.2.5.3 BPI Worst Pain

Missing baseline or post-baseline BPI worst pain score will not be imputed.

9.2.5.4 LAD

Subjects answering “no” to cutting down on things (Question 1) will be assigned 0 day of cutting down on things for Questions 2a (due to health reasons) and 2b (due to a fracture if a fracture is reported before the PRO assessment). For subjects answering

"subject does not know" to cut down on things (Question 1), the missing days of cutting down on things will not be imputed. For subjects answering "yes" to cut down on things (Question 1) but not knowing the number of days cutting down on things in Question 2a or Question 2b, the missing days of cutting down on things will not be imputed. Days of cutting down on things will not be imputed if the question is not answered at all.

Similarly, subjects answering "no" to hospitalization overnight (Question 3) will be assigned 0 day of hospitalization overnight for Question 3a (due to health reasons) and 3b (due to a fracture if a fracture is reported before the PRO assessment). For subjects answering "yes" to hospitalization overnight (Question 3) but not knowing the number of days of hospitalization overnight, the missing days will not be imputed. For subjects answering "subject does not know" to hospitalization overnight (Question 3), the missing days will not be imputed. Days of hospitalization overnight will not be imputed if the question is not answered at all. Similar approach will be applied to days of stay in bed (Question 4).

If the reported days of cutting down on things, hospitalization overnight, or stay in bed are greater than the recall period (30 days), minimum (30, reported days) will be applied to baseline and post-baseline visit.

9.2.6 Dates

No imputation will be done on incomplete stop date of an adverse event or a concomitant medication unless specified otherwise. For the purposes of deriving time-to-event variable from adverse event data or identifying from which point forward the data collected should be excluded or censored for the per-protocol analysis because of subjects receiving incorrect investigational product or proscribed therapies, incomplete start dates of adverse events or concomitant medications will be imputed as follows. Partial dates will be listed as is on the listings.

Table 3. Imputation Rules on Incomplete Start Date of Adverse Event or Concomitant Medication

	Missing	Impute	Exception
Start Date	Day	01	Default to Study Day 1 Date if the event started in the same year and month as **Study** Day 1
	Day / Month	01JAN	Default to Study Day 1 Date if the event started in the same year as **Study** Day 1
	Month	JAN	Default to Study Day 1 if the event started in the same year as Study Day 1
	Day / Month / Year	First Dose Date	Use randomization date for subjects who did not receive investigational product

If a death date (ie, End of Study date) is incomplete and missing only the day field, it will be imputed as the first day of the month if the latest date from other data is before the month of the death. However, if the latest assessment date is during the same month as the death, the partial death date will be imputed using the latest assessment date.

For dates of last period for menopause, the imputation rules for partial date are as follows: if the day is missing, default to day 15; if both month and day are missing, default to July 1. If the imputed date is on or after the randomization date, default to randomization date minus 1. Missing years will not be imputed.

If the date on the PRO administration page is partially or completely missing, the date of SC IP administration (or vital sign collection if SC IP is missing during the the double-blind period or during the open-label period) at the corresponding visit with the same CPEVENT will be used. Because it is expected that all PRO questionnaires be completed on the same day, EQ-5D-5L, BPI worst pain and LAD will use the same assessment date and the same imputation algorithm as that for OPAQ-SV.

9.2.7 Lab Parameters

Lab parameters with value below the lower limit of quantification (LLOQ) or above the upper limit of quantification (ULOQ) will be imputed as the LLOQ or ULOQ value, respectively.

9.2.8 Height

Missing post-baseline heights will be imputed using LOCF approach (by carrying forward the last nonmissing post-baseline value prior to the missing value). For the open-label period endpoints, only the open-label period value will be used for imputation, the values

in the double-blind period will not be carried forward into open-label period. The LOCF imputation will be done up to Month 24.

9.2.9 Oral IP Alendronate

Boxes containing oral IP alendronate/placebo are dispensed to subjects every 2 months during the double-blinded period and oral IP alendronate are dispensed every 6 months during the open-label period. The number of tablets returned are collected in eCRF. Before the datasnapshot, these data will be cleaned. If the number of returned alendronate were missing, it will be imputed that subjects did not take any IP.

9.3 Outliers

Scatter plots will be examined to identify potential outliers in any of the continuous variables identified in Section 4. Frequencies of the categorical data listed in Section 4 will be examined to identify questionable values. Before data snapshot the validity of any questionable values will be verified and observations found to be due to data entry errors will be corrected by the study team. Potential outliers that are not due to data entry error will be included in the analysis. No valid measurement will be purposely excluded from descriptive or inferential analyses. However, sensitivity analyses may be conducted to evaluate the influence of extreme values in the data. These analyses will be documented in the clinical study report.

9.4 Distributional Characteristics

The assumptions underlying the parametric models analyzed for continuous data will be checked. In cases where residuals indicate marked departures from the assumptions, additional sensitivity analyses will be performed using transformations or alternate methods such as nonparametric or robust procedures.

9.5 Validation and Configuration Management

Programs will be developed and maintained, and output will be verified according to processes described in procedures or technical manuals about the “Configuration Management of Statistical Analysis and Reporting Systems”, “Statistical Analysis and Reporting System Development and Validation”, and “Development of Statistical Analysis and Reporting Systems”.

Tables, figures and listings will be produced with validated standard macro programs where standard macros can produce the specified outputs.

The production environment consists of Amgen-supported versions of The SAS System running on the Sun Solaris operating system. Because it is common for multiple versions of SAS to be available during the study period, the SAS version used to produce analyses will be documented in the validation documentation and the clinical study report.

10. Statistical Methods of Analysis

10.1 General Principles

The analytical approach for this study will be to use inferential testing to evaluate the following:

- the effect of romosozumab treatment for 12 months followed by ALN treatment on fracture outcome as compared with ALN alone
- the effect of romosozumab treatment for 12 months on fracture outcomes as compared with ALN alone

All efficacy analyses will be performed by randomized treatment, regardless of actual treatment received.

At primary analysis, to maintain the overall significance level of 0.05, Hochberg's method will be used to evaluate the primary endpoints:

- Subject incidence of clinical fracture at primary analysis
- Subject incidence of new vertebral fracture through Month 24

If both the primary endpoints are significant at the 0.05 level (2-sided), each of the following secondary DXA BMD endpoints will be tested hierarchically at 0.05 (2-sided) according to the following sequence: percent change from baseline in BMD at lumbar spine at Month 24, percent change from baseline in BMD at total hip at Month 24, percent change from baseline in BMD at femoral neck at Month 24, percent change from baseline in BMD at lumbar spine at Month 12, percent change from baseline in BMD at total hip at Month 12, and percent change from baseline in BMD at femoral neck at Month 12.

If all preceding endpoints are significant, the nonvertebral fracture at the primary analysis will be evaluated based on a 1-sided test at the significance level determined by the alpha spending function specified in Section 3.2.3. The adjusted 2-sided p-value will be provided to demonstrate the overall significance level of 0.05. All remaining secondary and exploratory efficacy endpoints will be explored at significance level of

0.05 (2-sided). If superiority of the nonvertebral fracture endpoint is achieved at the primary analysis and the study is stopped after the primary analysis has been performed, all data, including the additional safety and nonvertebral fracture data, collected after the primary analysis will be summarized descriptively. Study completion constitutes either completion of primary analysis if superiority is proven for nonvertebral fractures or until 440 subjects experience a nonvertebral fracture.

For analyses where the age stratification factor needs to be incorporated, the following general principles will be followed:

- For stratified analyses that are intended to evaluate the treatment effect, the analyses will be done using the randomized stratum, regardless of the subject's actual stratum. If the stratification error rate is $> 5\%$, a sensitivity analysis using the stratification factor of the actual value will be done for the primary efficacy endpoint.
- When the stratification factor is used as a covariate in the covariate analysis where the analysis interest is on the association between the covariate and the outcome, the stratification factor based on the actual value will be used.

Continuous parameters will be summarized using descriptive statistics, which include the mean, standard deviation, minimum, 25th percentile, median (50th percentile), 75th percentile, maximum, and number of non-missing observations (n). Nominal and ordinal categorical variables will be summarized using frequencies and percentages. Unless otherwise specified, the percentages will be based on the number of subjects in the specific analysis set for the endpoint. Time-to-event parameters will be summarized using the Kaplan-Meier event rates.

10.2 Subject Accountability

The disposition of all randomized subjects will be tabulated by randomized treatment group. Subject enrollment and disposition for the number of subjects randomized, number of randomized subjects participating in the sub-study, successfully completing investigational product administration in the double-blind and overall study periods, and completing the study will be included. The disposition of subjects will also include the number of subjects who withdrew from the investigational product and their reasons for withdrawal for SC and oral investigation product. The disposition of the number of subjects who withdrew from study and their reasons for withdrawal will be provided.

10.3 Important Protocol Deviations

Important Protocol Deviations (IPDs) categories are defined by the study team before the first subject visit and updated during the IPD reviews throughout the study prior to database lock. These definitions of IPD categories, sub-category codes and descriptions will be used during the course of the study. The final IPD list is used to produce the Summary of IPDs table and the List of Subjects with IPDs. Important protocol violations including eligibility violations will be summarized and listed.

10.4 Demographic and Baseline Characteristics

The following demographic and baseline disease characteristics will be summarized by randomized treatment group based on all randomized subjects:

- Race
- Age (years)
- Age groups (<65, <75 and ≥ 75)
- Years since menopause
- Body composition (height [cm], weight [kg], and BMI [kg/m^2])
- 10-year probability of major osteoporotic and hip fractures based on WHO risk factor criteria (FRAX) calculated with femoral neck BMD T-score
- 10-year probability of major osteoporotic and hip fractures based on WHO risk factor criteria (FRAX) calculated without femoral neck BMD T-score
- Fracture history including any fracture, osteoporotic fractures, vertebral fractures, nonvertebral fractures, major nonvertebral fracture, any historical fracture after age 55, and any historical fracture after age 45
- Presence or absence of severe vertebral fracture at baseline; based on most severe Genant semiquantitative grade at baseline on screening spinal radiograph)
- Number of prevalent vertebral fractures at baseline (0, 1, 2, ≥ 3; based on screening spinal radiograph)
- Most severe Genant semiquantitative grade at baseline (based on screening spinal radiograph)
- BMD T-score of lumbar spine, total hip, and femoral neck

- Selected laboratory analytes (calcium corrected by albumin, phosphorus, creatinine, estimated GFR, and estimated GFR level [< 15, 15 to < 30, 30 to < 60, 60 to < 90, and ≥ 90 mL/min])
- Baseline serum 25 (OH) vitamin D level
- Prior use of any osteoporosis medication (yes, no)
    - Prior use of oral bisphosphonate (yes, no)
    - Prior use of intravenous bisphosphonate (yes, no)
    - Prior use of SERM (yes, no)
    - Prior use of PTH or PTH derivatives (yes, no)
    - Prior use of hormone replacement therapy (yes, no)
    - Prior use of calcitonin (yes, no)
    - Prior use of strontium (yes, no)
    - Prior use of fluoride (yes, no)
    - Prior use of calcitriol (yes, no)
    - Prior use of denosumab (yes, no)
- Baseline use of calcium and vitamin D (yes, no)
- Substance use during the past 5 years including tobacco use (never, former, and currently); alcoholic beverages (none, ≤ 2 per day, ≥ 3 per day);

10.4.1 Vitamin D Loading Dose Administration

Subjects with a serum 25 (OH) vitamin D level ≥ 20 ng/mL and ≤ 40 ng/mL at screening were to receive an initial loading dose of 50,000 to 60,000 IU vitamin D after randomization. Receipt of this loading dose will be summarized by baseline vitamin D level (< 20 ng/mL, ≥ 20 ng/mL and ≤ 40 ng/mL, and > 40 ng/mL) and randomized treatment group. Baseline vitamin D level is based on the last screening level prior to randomization. The loading dose will be identified based on the concomitant medication eCRF as follows:

1. Select vitamin D medication records with start date within 1 week of Study Day 1
2. Derive dose per day and total dose as:
    a. If dose ≥ 50,000 IU then dose per day and total dose are set to recorded dose

b. If dose < 50,000 IU then do the following to derive the total dose received:

i. If duration is missing or greater than 7 days, set duration to 7 days

ii. Dose per day and total dose are derived as follows based on recorded dose, frequency, and duration:

Frequency	Frequency code	Dose Per Day (DPD)	Total Dose
PER YEAR	PA	CMDOSE	CMDOSE
EVERY 2 MONTHS	Q2M	CMDOSE	CMDOSE
EVERY 2 WEEKS	Q2S	CMDOSE	CMDOSE
EVERY 3 MONTHS	Q3M	CMDOSE	CMDOSE
EVERY 3 WEEKS	Q3S	CMDOSE	CMDOSE
EVERY 4 WEEKS	Q4S	CMDOSE	CMDOSE
EVERY 6 MONTHS	Q6M	CMDOSE	CMDOSE
EVERY 6 WEEKS	Q6S	CMDOSE	CMDOSE
EVERY 8 MONTHS	Q8S	CMDOSE	CMDOSE
EVERY 8 WEEKS	Q8W	CMDOSE	CMDOSE
EVERY MONTH	QM	CMDOSE	CMDOSE
EVERY WEEK	QS	CMDOSE	CMDOSE
ONCE	ONCE	CMDOSE	CMDOSE
IMMEDIATELY	STAT	CMDOSE	CMDOSE
AT BEDTIME	HS	CMDOSE	DPD*duration
AS NEEDED	PRN	CMDOSE	DPD*duration
DAILY	QD	CMDOSE	DPD*duration
TWICE PER DAY	BID	CMDOSE * 2	DPD*duration
EVERY 12 HOURS	Q12H	CMDOSE * 2	DPD*duration
TWICE PER WEEK	BIS	CMDOSE * 2/7	DPD*duration
EVERY HOUR	QH	CMDOSE * 24	DPD*duration
EVERY 8 HOURS	Q8H	CMDOSE * 3	DPD*duration
3 TIMES PER DAY	TID	CMDOSE * 3	DPD*duration
THREE TIMES A WEEK	TIS	CMDOSE * 3/7	DPD*duration
EVERY 6 HOURS	Q6H	CMDOSE * 4	DPD*duration
4 TIMES PER DAY	QID	CMDOSE * 4	DPD*duration
4 TIMES PER WEEK	QIS	CMDOSE * 4/7	DPD*duration

Frequency	Frequency code	Dose Per Day (DPD)	Total Dose
EVERY 4 HOURS	Q4H	CMDOSE * 6	DPD*duration
EVERY OTHER DAY	QOD	CMDOSE / 2	DPD*duration
EVERY 3 DAYS	Q3D	CMDOSE / 3	DPD*duration
EVERY 4 DAYS	Q4D	CMDOSE / 4	DPD*duration

Page 2 of 2

3. Sum all total dose records for each subject to get total daily dose

4. Subjects with a total daily dose ≥ 50,000 IU will be considered as having received the loading dose of vitamin D

For subjects who enrolled in the sub-study, demographics, body composition, baseline BMD, and the following baseline laboratory data will also be summarized:

- Bone turnover markers (CTX, OC, BSAP, and P1NP)
- Selected laboratory analytes (calcium corrected by albumin, phosphorus, and iPTH)

10.5 Efficacy Analyses

The age stratification factor will be included in the statistical models when analyzing data from the entire study but will not be included in analyses of sub-study.

10.5.1 Descriptions of Standard Summaries and Statistical Models

10.5.1.1 Standard Analyses for Binary Endpoints

All binary endpoints will be summarized using the number and percentage of subjects having the response of interest by treatment group. Unless otherwise specified, the percentages will be based on the number of subjects in the specific analysis set for the endpoint.

Treatment comparisons on binary endpoints will be done using a logistic regression model (Agresti, 1990) with treatment as the main effect (using control arm as the reference category) and age stratification, baseline total hip BMD T-score, and presence or absence of severe vertebral fracture at baseline as covariate to assess the significance of the treatment difference between treatment arm and control arm. The odds ratio and the corresponding 95% 2-sided confidence interval and p-values using a

score test will be provided. In the case that the logistic regression model does not converge due to sparse data or separation, a conditional logistic regression stratified by age strata, and baseline total hip BMD T-score and presence or absence of severe vertebral fracture at baseline will be used. This will also apply to other analyses using logistic regression in this analysis plan (SAS code in Section 13.4.1)

In addition to the estimate of the odds ratio from the logistic regression model, point estimates of absolute risk reduction (difference in proportions, control – treatment) and risk ratio (ratio of proportions, treatment over control) as well as the corresponding 95% confidence intervals will also be calculated via the Mantel-Haenszel method (Agresti and Hartzel, 2000) adjusting for age strata, total hip BMD T-score ($\leq$ -2.5 and $>$ -2.5), and presence or absence of severe vertebral fracture at baseline (SAS code in Section 13.4.1, Section 13.4.2 and Section 13.4.3).

10.5.1.2 Standard Analyses for Ordinal Endpoints

All ordinal endpoints will be summarized using the number and percent of subjects in each category by treatment group. Unless otherwise specified, the percentages will be based on the subjects with non-missing observations. Treatment comparisons on ordinal endpoints will be done using a repeated measures proportional odds model adjusting for age strata and presence or absence of severe vertebral fracture at baseline, visit and treatment by visit interaction. The odds ratios and their corresponding 95% 2-sided Wald confidence interval will be provided (SAS code in Section 13.4.5).

10.5.1.3 Standard Analyses for Time-to-event Endpoints

All time-to-event endpoints will be summarized descriptively using the Kaplan-Meier estimates at time point(s) of interest (SAS code in Section 13.4.7). Further analysis using a stratified Cox proportional hazards model controlling for age strata with presence or absence of severe vertebral fracture at baseline, baseline total hip BMD T-score, and treatment as the independent variable will also be provided (SAS code in Section 13.4.8). The significance of the treatment effect between treatment arm and control arm will be assessed based on the Cox proportional hazards model. The estimated hazard ratio, corresponding 95% confidence interval, and the p-value of the score test from the model will be provided. In addition, the point estimate of the adjusted risk difference (difference in Kaplan-Meier estimates at the time point of interest, control arm - treatment arm) and the corresponding 95% confidence interval will be provided

using the inverse variance-weighted method (SAS code in Section 13.4.14). For 12 months analysis, subjects without having the event of interest in the double-blind period will be censored at the end of double-blind period date. For 24 months analysis, subjects without having the event of interest in the study will be censored at the end of 24-month study period.

The proportional hazards assumption will be examined visually using log-log survival plots versus time and plots of Schoenfeld residuals versus time (Schoenfeld, 1982). If the proportional hazards assumption is violated, a piecewise Cox proportional hazards model assuming proportional hazards in the 1st year, the 2nd year, and all remaining years thereafter, will be used.

10.5.1.4 Standard Analyses for Continuous Endpoints

All continuous endpoints will be summarized using descriptive statistics including mean, standard deviation, minimum, 25th percentile, median (50th percentile), 75th percentile, maximum, and number of nonmissing observations (n).

10.5.1.5 ANCOVA Model

The ANCOVA model (SAS code in Section 13.4.9) will include treatment, age (stratification factor), presence or absence of severe vertebral fracture at baseline, and baseline value of the endpoint. The model will allow different variances for the treatment groups for DXA or QCT endpoints. Additional covariates of machine type and machine type-by-baseline value interaction will also be included in the model to adjust for the effect of machine type on baseline value for parameters derived by DXA. The age strata will not be included in analyses for the sub-study.

The least-squares mean of the treatment difference (treatment – control) and the corresponding 95% confidence interval will be summarized by time points of interest, where each time point is estimated using a separate ANCOVA model. Graphical display of the least-squares means by randomized treatment group across time points may also be provided to visually summarize the results from the model.

10.5.1.6 Repeated Measures Model

The likelihood-based repeated measures model (Longford, 1993) will include treatment, age strata (stratification factor), presence or absence of severe vertebral fracture at baseline, visit, baseline value of the endpoint, and treatment-by-visit interaction as fixed

effects using an unstructured variance-covariance structure. The model will allow different variances for the treatment groups for endpoints. Additional covariates of machine type (Hologic or Lunar) and machine type-by-baseline value interaction will also be included in the model to adjust for the effect of machine type on baseline value for parameters derived by DXA. The visit will be treated as a categorical variable. The age strata will not be included in analyses for the sub-study. The approach of Kenward and Roger (1997) for estimating the denominator degrees of freedom for the hypothesis test will be followed (SAS code in Section 13.4.10). Other variance-covariance structure may be substituted if converg ence problem arises. The least-squares mean of the treatment difference (treatment – control) and the corresponding 95% confidence interval will be summarized by time points of interest. Graphical displays of least-squares means by randomized treatment group across time points may also be provided to visually summarize the results from the model.

10.5.1.7 Wilcoxon Rank-sum Test

Wilcoxon rank-sum test (SAS code in Section 13.4.11) will be used to assess the significance of the treatment difference at each time point for bone turnover markers and biomarkers and any other endpoints where parametric methods may not be appropriate.

10.5.1.8 Van Elteren Stratified Rank Test

Van Elteren stratified rank test (van Elteren, 1960; SAS code in Section 13.4.13) adjusting for age strata (stratification factor) and presence of severe vertebral fracture at baseline will be used to assess the significance of the treatment difference at each time point for continuous data from the main study where parametric methods may not be appropriate.

10.5.2 Analyses of Primary Efficacy Endpoints

10.5.2.1 Primary Analytical Approach

The primary analytical model for comparing subject incidence of new vertebral fractures through Month 24 between the randomized treatment groups will use a logistic regression model as described in Section 10.5.1.1 based on the primary efficacy analysis set for vertebral fractures as described in Section 7.1. The number and percentage of subjects with ≥ 1 new vertebral fracture through Month 24 will be summarized by randomized treatment group. The adjusted odds ratio of romosozumab

as compared with ALN and the corresponding 95% confidence interval will also be provided.

The primary analytical model for comparing subject incidence of clinical fractures at primary analysis between the randomized treatment groups will use a stratified Cox proportional hazards model as described in Section 10.5.1.3 based on the full analysis set as described in Section 7.2. The cumulative incidence of fractures will be summarized using the Kaplan-Meier estimates, and the hazard ratios of romosozumab as compared with ALN and the confidence intervals for the risk reduction relative to ALN control arm will be based on the model.

10.5.2.2 Supportive Analyses

To demonstrate the robustness of the results from the primary analytical models, additional supportive analyses will be performed:

Per protocol analyses

Primary analytical models will be repeated using the per protocol subset as described in Section 7.4 for new vertebral fracture through Month 24 and clinical fracture at primary analysis. For subjects who received the investigational product not matching their randomized treatment group or proscribed therapy on study and had investigational product IPD, their vertebral fracture evaluations will be ignored at or after the first occurrence of violation. The imputation described in Section 9.2.2 will be used to impute their new vertebral fracture status at Month 24.

Time-to-event analysis based on full analysis set

A supportive analysis for new vertebral fracture through Month 24 will be performed based on time-to-first new vertebral fracture using the stratified Cox proportional hazards model as described in Section 10.5.1.3.

10.5.2.3 Covariate Analysis

For new vertebral fracture the odds ratio of romosozumab as compared with ALN and corresponding 95% confidence interval and p-value, adjusting for each of the covariates listed in Section 4.6 using the logistic regression model as described in Section 10.5.1.1, as well as adjusting for all covariates simultaneously in a multivariate analysis, will be presented. The analysis subset and the imputation method will be the same as those used in the primary analytical model.

Similarly, for clinical fracture, the adjusted hazard ratio of romosozumab as compared with ALN and corresponding 95% confidence interval and p-value adjusting for each of the covariates listed in Section 4.6 using the stratified Cox proportional hazards model as described in Section 10.5.1.3, as well as adjusting for all covariates simultaneously in a multiple regression analysis, will be presented.

10.5.2.4 Subgroup Analysis

To assess the consistency of the treatment effect in different subgroups, the incidence of new vertebral fractures and incidence of clinical fracture will be analyzed within each subgroup stratum listed in Section 7.9. For incidence of new vertebral fracture, the odds ratio of romosozumab as compared with ALN, and their respective 95% confidence intervals, will also be summarized for each category of a subgroup. The treatment-by-subgroup interaction will be tested using a logistic regression as described in Section 10.5.1.1 and including individual subgroup and treatment-by-subgroup interaction. If the p-value of the interaction is ≥ 0.05, the treatment-by-subgroup interaction is considered not significant. Otherwise, a 2-sided Gail and Simon test (Gail and Simon, 1985) will be used to test whether there is qualitative interaction (SAS code in Section 13.4.6).

For clinical fracture, the hazard ratio of romosozumab as compared with ALN, and their respective 95% confidence intervals, will also be summarized for each category of a subgroup. The treatment-by-subgroup interaction will be tested using stratified Cox proportional hazards model as described in Section 10.5.1.3 and individual subgroup and treatment-by-subgroup interaction. Similarly, the interaction will be further evaluated for qualitative interaction using Gail and Simon test if the p-value of the score test for the interaction is ≥ 0.05.

The verified age will be used to stratify subgroups for the analysis of age subgroup.

10.5.3 Analyses of Secondary and Exploratory Efficacy Endpoints

10.5.3.1 Nonvertebral Fracture, Clinical Fracture, Clinical Vertebral Fracture, All Fracture, Major Nonvertebral Fracture, Major Osteoporotic Fracture and Hip Fracture

The full analysis set as described in Section 7.2 will be used for the analyses of following efficacy endpoints:

- subject incidence of nonvertebral fracture through Month 12, through Month 24, at primary analysis and final analysis
- subject incidence of clinical fracture through Month 12 and through Month 24

- subject incidence of clinical vertebral fracture through Month 12 and through Month 24
- subject incidence of all fracture through Month 12 and at primary analysis
- subject incidence of major nonvertebral fracture through Month 12, at primary analysis and at final analysis
- subject incidence of major osteoporotic fracture through Month 12 and at primary analysis
- subject incidence of hip fracture through Month 12, through Month 24, at primary analysis and at final analysis.

As described in Section 10.5.1.3, the cumulative incidence of fractures will be summarized using the Kaplan-Meier estimates, and the hazard ratios of romosozumab as compared with ALN and the confidence intervals will be based on the stratified Cox proportional hazards model.

The primary analytical model for clinical fracture through Month 12, nonvertebral fracture through Month 12 ,at primary and final analyses, will be repeated using the per protocol subset as described in Section 7.3 and 7.4 as a supportive analysis. Further, the hazard ratio and corresponding 95% confidence interval and p-value adjusting for each of the covariates listed in Section 4.6 using the stratified Cox proportional hazards model as described in Section 10.5.1.3, as well as adjusting for all covariates simultaneously in a multiple regression analysis, will be presented.

For nonvertebral fracture through Month 12 and at primary analysis, the hazard ratio of romosozumab as compared with ALN, and their respective 95% confidence intervals, will also be summarized for each category of a subgroup. The treatment-by-subgroup interaction will be tested using stratified Cox proportional hazards model as described in Section 10.5.1.3 and individual subgroup and treatment-by-subgroup interaction. Similarly, the interaction will be further evaluated for qualitative interaction using Gail and Simon test if the p-value of the score test for the interaction is ≥ 0.05.

The efficacy analysis set in Month 12 to 24 study period as described in Section 7.10 will be used for the analyses of following efficacy endpoints between Month 12 and Month 24: subject incidence of clinical fracture, subject incidence of nonvertebral fracture, subject incidence of hip fracture, and subject incidence of clinical vertebral fracture. The incidence of fractures will be summarized using the Kaplan-Meier estimates, and the hazard ratios of romosozumab as compared with ALN and the confidence intervals will be based on the stratified Cox proportional hazards model as described in

Section 10.5.1.3. Due to the lack of randomization at Month 12, the comparability between the two arms is not ensured and therefore the analysis is more susceptible to bias.

10.5.3.2 New Vertebral Fracture, New or Worsening Vertebral Fracture, and Multiple New or Worsening Vertebral Fracture

The primary efficacy analysis set for vertebral fracture as described in Section 7.1 will be used for the analyses of following efficacy endpoints: subject incidence of new vertebral fracture through Month 12, subject incidence of new or worsening vertebral fracture through Month 12 and Month 24, subject incidence of multiple new or worsening vertebral fracture through Month 12 and Month 24. The above subject incidences between subjects in the treatment group and subjects in the control group will be compared using a logistic regression model adjusting for age strata as described in Section 10.5.1.1. The number and percentage of subjects with ≥ 1 above specified vertebral fracture will be summarized by treatment group for first 12-month of double-blind period and for 24-month study period. The adjusted odds ratio and the corresponding 95% confidence interval will also be provided.

The efficacy analysis set for vertebral fracture during 24-month study period as described in Section 7.11 will be used for the analysis of subject incidence of new vertebral fractures between Month 12 and Month 24. The number and percent of subjects with new vertebral fracture between Month 12 and Month 24 in this efficacy analysis set will be summarized by treatment group. The treatment comparison will be done using a logistic regression model as described in Section 10.5.1.1. Due to the lack of randomization at Month 12, the comparability between the two arms is not ensured and therefore the analysis is more susceptible to bias.

10.5.3.3 Percent Changes in Lumbar Spine, Total Hip and Femoral Neck BMD by DXA

The primary efficacy subset for BMD described in Section 7.6 will be used for the analyses of percent changes from baseline in lumbar spine, total hip and femoral neck by DXA at month 12**,** 24**,** and 36.

The treatment comparisons of the BMD in lumbar spine, total hip and femoral neck at months 12 and 24 will be analyzed using the ANCOVA model described in Section 10.5.1.5 with LOCF imputation as the primary analysis. Descriptive summary of

observed DXA values at Month 36 will be provided without LOCF. Graphical displays to assess the model will include response versus fitted values, response versus standardized residuals and a normal probability plot of the standardized residuals. The sensitivity analyses include the repeated measures model as described in Section 10.5.1.6, where each body site is estimated using a separate repeated measures model.

Absolute values and change from baseline for BMD will also be summarized descriptively by time point and machine type.

10.5.3.4 Change from Baseline in Height

The subset as described in Section 7.6 will be used for the analyses of change from baseline in height at months 12 and 24. A nonparametric method, Van Elteren rank test, as described in Section 10.5.1.8, will be used for treatment comparisons for change from baseline in height. ANCOVA models as described in Section 10.5.1.5 will also be used for treatment comparisons.

10.5.4 PRO/ClinRO Exploratory Endpoints

The subset described in Section 7.6 will be used for the analysis of PRO/ClinRO endpoints and the subset described in Section 7.7 will be used for the analysis of post fracture PRO endpoints.

Descriptive summary statistics will be provided for actual values and changes from study baseline by visit as described in Section 10.5.1.4. Descriptive summary statistics will also be provided for the changes from the pre-fracture baseline by visit since fracture. The comparison of treatments on the post-fracture PRO data includes subjects who have a clinical fracture on study, however, the fracture type, fracture location and patient characteristics might be different in these fractured subjects between two treatment arms. Although various factors will be adjusted in the analysis, due to the lack of randomization, the comparability is not ensured and therefore the analysis is more susceptible to bias.

OPAQ-SV

The analyses of changes from study baseline at months 6, 12, 18, 24, 30, and 36 in OPAQ-SV dimension scores of physical function, emotional status, and back pain will be performed using the repeated measures model separately as described in

Section 10.5.1.6. The analyses of changes from pre-fracture baseline at month 1, 2 and 3 after reporting a clinical fracture in OPAQ-SV dimension scores will be performed using the repeated measures model adjusting for age strata, presence or absence of severe vertebral fracture at baseline, treatment, visit, fracture type, pre-fracture baseline score, change from pre-fracture baseline at report of fracture and treatment by visit interaction. The missing OPAQ SV dimension scores at reporting of the fracture will be handled by multiple imputation.

EQ-5D-5L

The analyses of changes from study baseline at months 6, 12, 18, 24, 30, and 36 in EQ-5D-5L VAS score will be performed using the repeated measures model as described in Section 10.5.1.6.

Each of the 5 EQ-5D-5L dimension scores (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) at month 6, 12, 18, 24, 30, and 36 will be analyzed using repeated measure proportional odds model adjusting for age strata, presence or absence of severe vertebral fracture at baseline, treatment, visit, baseline score and treatment-by-visit interaction as described in Section 10.5.1.2.

The change from study baseline at month 6, 12, 18 24, 30, 36 and the change from pre-fracture baseline at month 1, 2 and 3 after fracture in 5 EQ-5D-5L dimension scores will be summarized by shift tables.

The analyses of changes from pre-fracture baseline at month 1, 2 and 3 after reporting a clinical fracture in EQ-5D-5L VAS score will be performed using the repeated measures model adjusting for age strata, presence or absence of severe vertebral fracture at baseline, treatment, visit, fracture type, pre-fracture baseline score, change from pre-fracture baseline at report of fracture and treatment by visit interaction. The missing EQ-5D-5L VAS score at reporting of the fracture will be handled by multiple imputation.

BPI Worst Pain Score

The analyses of changes from study baseline at months 6, 12, 18, 24, 30, and 36 in BPI worst pain score will be performed using the repeated measures model as described in Section 10.5.1.6. The analyses of changes from pre-fracture baseline at month 1, 2 and 3 after reporting a clinical fracture in BPI worst pain score will be performed using the repeated measures model adjusting for age strata, presence or absence of severe

vertebral fracture at baseline, treatment, visit, fracture type, pre-fracture baseline score, change from pre-fracture baseline at report of fracture and treatment by visit interaction. The missing BPI worst pain score at reporting of the fracture will be handled by multiple imputation.

The proportion of subjects with a clinically meaningful improvement in worst pain in 3 months after experiencing a nonvertebral or clinical vertebral fracture (defined as a 2-point improvement in the BPI worst pain scale compared with the fracture reporting visit) will be analyzed using a logistic regression model adjusting for age strata, presence or absence of severe vertebral fracture at baseline, treatment, fracture type, worst pain score at report of fracture and change from pre-fracture baseline at report of fracture. The proportion of subjects with a clinically meaningful improvement in worst pain at each post fracture reporting visit and within 3 months from pre-fracture baseline will be analyzed using same approach.

LAD Questionnaire

The number and percent of subjects reporting in the past month cutting down on things due to health reasons (Question 1), in hospital overnight due to health reasons (Question 3), and stay in bed due to health reasons (Question 4) at months 6, 12, 18, 24, 30, and 36, and at month 1, 2 and 3 after fracture will be summarized descriptively. Subjects reporting cutting down things but with missing value for number of days will not be included in the respective analysis of number of days.

Number of days in the past month of cutting down things due to health reasons (Question 2a), in hospital overnight due to health reasons (Question 3a), and stay in bed due to health reasons (Question 4a) at months 6, 12, 18, 24, 30, and 36 will be summarized descriptively. The treatment comparison at months 6, 12, 18, and 24, will be done using the repeated measures model adjusted for age strata, presence or absence of severe vertebral fracture at baseline, treatment, visit, treatment by visit interaction, and baseline.

Number of days in the past month of cutting down on things due to a fracture (Question 2b), in hospital overnight due to a fracture (Question 3b), and stay in bed due to a fracture (Question 4b) at month 1, month 2 and month 3 after fracture will be summarized descriptively in subjects who have a fracture during the 12-month double-blind period. The treatment comparison will be done using the repeated

measures model adjusted for age strata, presence or absence of severe vertebral fracture at baseline, treatment, visit, treatment by visit interaction, fracture type, pre-fracture baseline and change from pre-fracture baseline at report of fracture.

Change from pre-fracture baseline in number of days in the past month of cutting down things due to health reasons, in hospital overnight due to health reasons, and stay in bed due to health reasons at month 1, month 2 and month 3 after fracture and each first post fracture visit will be summarized descriptively in subjects who have a fracture during the 12-month double-blind period. The treatment comparison will be done using the repeated measures model adjusted for age strata, presence or absence of severe vertebral fracture at baseline, treatment, visit, fracture type, pre-fracture baseline, change from pre-fracture baseline at report of fracture and treatment by visit interaction.

Change from pre-fracture baseline in number of days in the past month of cutting down things due to a fracture, in hospital overnight due to a fracture, and stay in bed due to a fracture at fracture reporting visit and each first post fracture visit will be summarized descriptively in subjects who have a fracture during the 12-month double-blind period. The treatment comparison will be done using the repeated measures model adjusted for age strata, presence or absence of severe vertebral fracture at baseline, treatment, visit, fracture type, pre-fracture baseline, change from pre-fracture baseline at report of fracture and treatment by visit interaction.

10.6 Safety Analyses

Safety data will be summarized for subjects who received ≥ 1 dose of active investigational product by the actual treatment received in the double-blind period (any subject randomized to ALN arm who incorrectly receives ≥ 1 dose of romosozumab will be analyzed as receiving romosozumab).

Safety data for the double-blind period and for primary analysis period will be summarized separately using analysis set defined in Section 7.5. Safety data will also be summarized for the overall study period (at final analysis) using analysis set defined in Section 7.5.

A listing of adverse events will be prepared for subjects who did not received any active dose of investigational product in the 12-month of double-blind period. A listing of adverse events occurring after double-blind period will be prepared for subjects

who received at least one **active** dose of investigational product in the 12-month of double-blind period, remained on study in the open-label period without open-label alendronate exposure.

10.6.1 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) version 19.1 or later will be used to code all adverse events to a system organ class and a preferred term for the 12-month double-blind period, the primary analysis period (at the primary analysis) , and the overall study period (at the final analysis). Adverse events will be summarized in the double-blind period, in the primary analysis study period (at the primary analysis), and overall study period (at the final analysis) for all subjects who receive at least one dose of investigational product in the double-blind period. The subject incidence rates for the primary study and overall study period will include all events that occurred in the double-blind period, and in addition, all events that occurred in the open-label period for those subjects who received at least one dose of open-label ALN.

All adverse event tables will be summarized by actual treatment group. The subject incidence of adverse events will be summarized for all treatment-emergent, serious, treatment-related, serious treatment-related, those leading to withdrawal of study, leading to withdrawal of investigational product, fatal, and of special interest. Subject incidence of special interest adverse events will also be summarized according to their categories.

Subject incidence of all treatment-emergent, serious, treatment-related, serious treatment-related, those leading to withdrawal of study, leading to withdrawal of investigational product, and fatal adverse events will be tabulated by system organ class and preferred term in descending order of frequency. Subject incidence of all treatment-emergent adverse events will be tabulated by system organ class, high-level group term, and preferred term in descending order of frequency as well as by preferred term.

Summaries of treatment-emergent adverse events occurring in at least 5% of the subjects and serious treatment-related adverse events occurring in at least 0.1% of the subjects will be provided by preferred term in descending order of frequency. Treatment-emergent adverse events occurring in at least 2% of the subjects and also

have a >1% difference than ALN will also be summarized by preferred term in descending order of frequency.

Subgroups group analyses for age strata (< 75, ≥ 75 years) will be presented by system organ class and preferred term in descending order of frequency.

In addition, subject incidence rates of adverse events corresponding to preferred terms of events of interest of hypocalcemia, injection site reactions, potentially related to hypersensitivity (Standardized MedDRA Query [SMQ], narrow scope), malignant or unspecified tumor (SMQ, narrow scope), hyperostosis, osteoarthritis, adjudicated adverse events of osteonecrosis of the jaw, and atypical femoral fracture will be provided.

Unless otherwise specified, there is no planned statistical testing in the safety analyses.

10.6.2 Osteonecrosis of the Jaw Adjudication

The events of osteonecrosis of the jaw (ONJ) which occurred in the study will be adjudicated and summarized. All potential events of ONJ identified through a pre-defined search of the MedDRA terms will be submitted to the Osteonecrosis of the Jaw Adjudication Committee for blinded review and adjudication. The committee will determine whether the event meets the case definition criteria for ONJ.

Number of adjudicated positive ONJ events will be summarized.

10.6.3 Adjudicated Positive Cardiovascular Events

All deaths and potential cardiovascular-related serious adverse events will be submitted to an external independent committee comprised of experienced cardiologists for blinded adjudication. The committee will adjudicate the events and determine whether the event is cardiovascular in nature.

Baseline cardiovascular risk factors will be summarized descriptively. These risk factors are defined as age (≥ 75, < 75), smoking history, and other clinical history including hypertension, diabetes, cardiovascular disease, and hypercholesterolemia. The cardiovascular risk factors of history of hypertension, diabetes, cardiovascular disease, and hypercholesterolemia will be identified based on the Medical & Surgical History eCRF based on MedDRA version 19.1. The SMQ for Hypertension, the SMQ for Hyperglycaemia, and the System Organ Classes of Cardiac Disorders and Vascular

Disorders will be used to identify preferred terms associated with hypertension, diabetes and cardiovascular disease, respectively. The Dyslipidaemia SMQ will be used to identify history of hypercholesterolemia, excluding terms associated with conditions of low cholesterol or increased high-density lipoprotein.

Only events confirmed positive by the adjudication committee to meet cardiovascular event definition criteria will be included for analyses. Adjudicated cardiovascular events of death, cardiac ischemic event, cerebrovascular event, non-coronary revascularization, heart failure and peripheral vascular events not requiring revascularization will be summarized using subject incidence rates, odds ratios and 95% confidence intervals. No statistical tests will be performed.

10.6.4 Atypical Femoral Fracture Adjudication

The events of atypical femoral fracture (AFF) which occurred in the study will be adjudicated and summarized. All potential events of AFF identified through a pre-defined search of the MedDRA terms will be submitted to the Atypical Femoral Fracture Adjudication Committee for blinded review and adjudication. The committee will determine whether the event meets the case definition criteria for AFF.

Number of adjudicated positive AFF events will be summarized.

10.6.5 On Study Cancer

Number and percentage of subjects who reported new malignancy on study will be summarized by cancer category and cancer status.

10.6.6 Injection Site Reaction

The number of episodes of injection site reaction, duration, nature (concomitant versus recurrent) and severity will be summarized during the double-blind study period. The time (in days) to first injection site reaction during the double-blind study period will also be summarized descriptively.

10.6.7 Hypersensitivity

Subject incidence of hypersensitivity as an EOI with on-set day within 2 days since receiving any dosing of investigational product by preferred term will be provided for the double-blind study period. Similarly, the analyses will be repeated for hypersensitivity with on-set day $>$ 2 and $\leq$ 7 days since receiving any dosing of investigational product.

In addition, subject incidence of concomitant/recurrent hypersensitivity EOIs, hypersensitivity leading to investigational product discontinuation, and hypersensitivity leading to study discontinuation will be provided by preferred term for the double-blind study period.

10.6.8 Tumor Necrosis Factor Mediated Inflammatory Diseases

Subject incidence rates of adverse events corresponding to a Amgen search strategy of preferred terms for tumor necrosis factor (TNF) mediated inflammatory diseases will be provided using the safety analysis set for all subjects, subjects with prior history of TNF disease, and subjects without prior history of TNF disease.

10.6.9 Laboratory Test Results

Subjects with missing data for a scheduled visit will not contribute to the safety tabulations for that time point (no imputation). Descriptive statistics for the actual value at each visit and the change from baseline at each post-baseline visit during the study period will be summarized. Figures showing central tendency and dispersion of the actual values and/or changes from baseline of select laboratory parameters (calcium corrected by albumin, phosphorus, magnesium and alkaline phosphatase) during the double-blind period will be provided. Above summary statistics and figures will also be done for percent change from baseline in calcium corrected by albumin, phosphorus, magnesium, and alkaline phosphatase. Visit windows will be used for these summaries as described in Section 13.1.1

Laboratory parameters will be summarized showing the shift from baseline to the most extreme post-baseline value between baseline and the end of double-blind period based on Common Terminology Criteria for Adverse Events v3.0 (CTCAE). Subject incidence of most extreme post-baseline calcium corrected by albumin decrease grade will also be summarized. The percentages of subjects with laboratory toxicities $\geq$ grade 2 CTCAE will be summarized. In addition, subject listings of grades 3 and 4 laboratory values will be provided.

Drug-induced liver injury will be assessed by evaluating subjects for Hy's Law criteria in the double-blind period and in the primary analysis study period. Hy's law lab value criteria is defined as aspartate amino transferase or alanine amino transferase $>$ 3 times upper limit of normal (ULN), total bilirubin $>$ 2 times ULN, and alkaline phosphatase $<$ 2 times ULN assessed within 7 days. Subjects who meet these lab criteria on study will

be further evaluated to assess whether there exist underlying conditions or concomitant medications which can explain the elevation in laboratory analytes.

Summaries of abnormal laboratory values, including summaries of shifts, worst post-baseline calcium, and evaluation of potential Hy's law cases will be summarized in the double-blind period and in the primary analysis period for all subjects who receive at least one dose of investigational product in the double-blind period. The subject incidence rates for the 24-month study period will include all laboratory measurements collected in the double-blind period, and in addition, all laboratory measurements collected in the open-label period for those subjects who received at least one dose of open-label alendronate.

10.6.10 Vital Signs, Height, Body Weight, and BMI

Descriptive statistics of actual values and changes from baseline in vital signs (systolic and diastolic blood pressure, heart rate, and temperature), body weight, and BMI will be presented by scheduled visit. Height, which is analyzed as an efficacy endpoint, will be summarized as well using the safety analysis set.

10.6.11 Anti-Romosozumab Antibody

Immunogenic response during the study will be described by tabulating the numbers and percentages of subjects who tested positive (binding and neutralizing) for anti-romosozumab antibodies in subjects receiving ≥ 1 dose of romosozumab at each visit and at any visit during study. Analysis will be performed using safety analysis set defined in Section 7.5. Subjects who test positive for either binding or neutralizing antibodies against romosozumab will be interpreted as persistently positive if the antibody status remains positive throughout the last time point tested. Subjects who test positive for binding antibodies against romosozumab will be interpreted as transient positive if the binding antibody status was negative at the subject's last time point tested within the study period. Subjects who test positive for neutralizing antibodies against romosozumab will be interpreted as transient positive if the neutralizing antibody status was negative at the subject's last time point tested within study period.

If a subject tests positive for antibodies against romosozumab, the relationship between the presence of antibodies, adverse events, concomitant medications, and bone mineral density will be evaluated; the immunogenic response will be listed by subject and study day along with the serum romosozumab levels if available. Note that the serum

romosozumab levels are only collected on subjects who participate in the sub-study. The subject incidence of injection site reaction EOIs, hypersensitivity EOIs, and adverse events corresponding to the MedDRA high level group term of autoimmune disorders will be provided by binding and neutralizing anti-romosozumab antibody status.

10.6.12 Exposure to Investigational Product

Descriptive statistics will be produced to describe the exposure of SC and oral investigational product by randomized treatment group for subjects exposed to investigational product for the double-blind period and open-label period, respectively. The total ALN exposure for the subject randomized to ALN arm will also be provided.

10.6.13 Exposure to Concomitant Medication

Types of concomitant medications will not be summarized.

10.7 Endpoints From Sub-study

10.7.1 Bone Turnover Marker and Biomarker

The subset described in Section 7.8.2 and 7.8.3 will be used for the analyses of percent change in bone turnover markers and biomarkers. These include values, changes from baseline, and percent changes from baseline at month 1, 3, 6, ,9, 12, 15, 18, 24, and 36 in P1NP, sCTX, BSAP, OC and sclerostin. The Wilcoxon rank-sum test as described in Section 10.5.1.7 will be used for treatment comparisons for percent change in bone turnover markers or biomarkers. In addition, percent changes from baseline at post-baseline visits in iPTH and serum sclerostin level will be provided.

10.7.2 BMD by DXA

The subset described in Section 7.8.1 will be used for the analyses of percent changes in lumbar spine, total hip and femoral neck BMD by DXA. Percent changes from baseline at months 6, 12, 18, and 24 in DXA BMD will be compared between treatment groups using a repeated measures model as described in Section 10.5.1.2 and an ANCOVA model described in Section 10.5.1.5 with treatment group, baseline BMD, machine type and interaction of baseline BMD and machine type as independent variables. Any missing value during the 12-month double-blind ALN-controlled study period will be imputed by carrying forward the last non-missing post-baseline value prior to the missing value. Descriptive summary of observed DXA values at Month 36 will be provided.

10.7.3 vBMD by QCT and Lumber Spine Strength by FEA

The primary efficacy subset described in Section 7.8.1 will be used for the analyses of percent changes from baseline in for integral (total) and trabecular volumetric BMD by QCT and lumbar spine strength by FEA. Percent changes from baseline at months 6, 12, and 24 in QCT vBMD and lumbar spine strength will be compared between treatment groups using a repeated measures model as described in Section 10.5.1.2 and an ANCOVA model described in Section 10.5.1.5 with treatment group and baseline vBMD as independent variables. Any missing value during the 12-month double-blind ALN-controlled study period will be imputed by carrying forward the last non-missing post-baseline value prior to the missing value.

10.7.4 PK

Serum concentration data described in Section 7.8.4 will be provided to PKDM group for analyses. Details regarding objectives, data handling, and methodology of the analysis will be provided in a separate analysis plan.

11. Changes From Protocol-Specified Analyses

Protocol Section 10.5.2.1 specified a logistic regression model adjusting for fracture location and potentially other factors impacting the pain score to analyze the incidence of subjects with a clinical meaningful improvement in worset pain. Fracture type is used as a covariate instead of the fracture location.

12. Literature Citations / References

Agresti A. Categorical Data Analysis, New York: John Wiley & Sons, Inc, 1990.

Agresti A and Hartzel J. Strategies for comparing treatments on a binary response with multi-centre data. Statistics in Medicine. 2000; 19:1115-1139.

Black DM, Cummings SR, Karpf DB, et al. Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. Fracture Intervention Trial Research Group. Lancet. 1996;348:1535-1541.

Black DM, Thompson DE, Bauer DC et al. Fracture Risk Reduction with Alendronate in Women with Osteoporosis: The Fracture Intervention Trial. J Clin Endocrinol Metab. 2000;85: 4118–4124.

Cummings SR, San Martin J, McClung MR, et al. Denosumab for prevention of fractures in postmenopausal women with osteoporosis. N Engl J Med. 2009;361(8):756-765.

Gail MH and Simon R. Testing for qualitative interactions between treatment effects and patients subsets. *Biometrics.* 1985; 41: 361-372.

Genant HK, Wu CY, Van Kuijk C, and Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. *Journal of Bone and Mineral Research.* 1993; 8(9): 1137-1148

Greenspan SL, Shneider DL, McClung MR, et al. Alendronate improves bone mineral density in elderly women with osteoporosis residing in long-term care facilities. A randomized, double-blind, placebo-controlled trail. *Ann Intern Med* 2002; 136:742-746

Hochberg Y. A Sharper Bonferroni Procedure for Multiple Tests of Significance. Biometrika. 1988; (74):800-802.

Kanis JA, Johnell O, Oden A, De Laet C, Johnsson B. Ten year probabilities of osteoporotic fractures according to BMD and diagnostic thresholds. *Osteoporosis International.* 2001; 12: 989-995.

Kanis JA, Johnell O, Oden A, Johansson H, and McCloskey E. FRAX™ and the assessment of fracture probability in men and women from the UK. *Osteoporosis International.* 2008; 19: 385-397.

Kenward MG and Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics. 1997; 53: 983-997

Longford NT. Random Coefficient Models, New York: Oxford University Press, Inc, 1993

Schnitzer T, Bone HG, Crepaldi G, et al. Therapeutic equivalence of alendronate 70 mg once weekly and alendronate 10 mg daily in the treatment of osteoporosis *Aging Clin Exp Res*. 2000; 25(5): 976-982.

Schoenfeld D. Partial Residuals for the Proportional Hazards Regression Model. *Biometrika*. 1982; 69: 239 -241.

Silverman SL. The osteoporosis assessment questionnaire (OPAQ): a reliable and valid disease-targeted measure of health-related quality of life (HRQOL) in osteoporosis. Quality of Life Research. 2000; 9: 767-774.

Van Elteren PH. On the combination of independent two-sample test of Wilcoxon, *Bulletin of the International Statistical Institute.* 1960; 37: 351-361.

13. Appendices

13.1 Technical Detail and Supplemental Information Regarding Statistical Procedures and Programs

13.1.1 Visit Windows

Per protocol, all monthly study visits up to Month 13 have a ± 7-day window. Following the 12-month double-blind ALN-controlled study period, study visits have a ± 14-day window. To allow for variations in scheduling, the following sets of visit windows will be used to assign evaluations to a most appropriate nominal visit for analysis. Furthermore, there will be no gaps between visit windows in order to include as many data points as possible for summarization.

Regardless of the width of the visit window, if more than 1 visit falls within the defined window, the result from the visit closest to the target day will be used. If 2 evaluations are of the same distance from the target day, the result from the later visit will be used. If more than one evaluation on the same date and time, the average of the results will be used. Only laboratory results collected from the central laboratory will be averaged in the case of duplicate results. Results not included for analysis will be included in listings only.

For LOCF imputation, the last evaluation on or prior to the upper bound of the specified window (ie, no lower bound) would be used for each nominal visit.

The window for the monthly phone contact in between clinic visits throughout the study is ± 7 days. The window for the end of primary analysis period phone call is −14 days to end of primary analysis. Unscheduled follow-up clinic visits after the monthly phone contact have a + 21-day window. The window for the end of study phone contact is −14 days to + 7 days.

13.1.1.1 Spine X-ray

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline [a]	1	Last evaluation prior to or on Study Day 1
Month 12	366	Study Day 2 to End of Double-blind Period Date
Month 24	731	End of Double-blind Period Date +**1** to End of Open-label M24 Study Period Date
Month 36	1096	End of Open-label M24 Study Period Date +**1** to Study Day 1278
Month 48	1461	Study Day 1279 to 1643
Month 60	1826	Study Day 1644 to 2008

[a] If results from baseline x-ray are not available, the results from scan taken on or before Study Day 14 will be considered baseline values.

Lateral spine X-rays are collected every 12 months after M36 until the end of the primary analysis period.

13.1.1.2 DXA (lumbar spine and proximal femur) for All Subjects

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline [a]	**1**	**Last evaluation prior to or on Study Day 1**
Month 12	**366**	**Study Day 2 to End of Double-blind Period Date + 30**
Month 24	**731**	**End of Double-blind Period Date + 31 to End of Open-label M24 Study Period Date + 30**
Month 36	**1096**	**End of Open-label M24 Study Period Date + 31 to Study Day 1278**
Month 48	**1461**	**Study Day 1279 to 1643**
Month 60	**1826**	**Study Day 1644 to 2008**

[a] If results from baseline DXA are not available, the results from scan taken on or before Study Day 14 will be considered baseline values and not the Month 12 values.

DXA of lumbar spine and proximal femur are collected every 12 months after M36 until the end of the primary analysis period.

13.1.1.3 DXA (lumbar spine and proximal femur) for imaging component of Sub-study

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline [a]	1	Last evaluation prior to or on Study Day 1
Month 6	183	Study Day 2 to 274
Month 12	366	Study Day 275 to End of Double-blind Period Date + **30**
Month 18	548	End of Double-blind Period Date +**31** to Study Day 639
Month 24	731	Study Day 640 to End of Open-label M24 Study Period Date +**30**
Month 36	1096	≥ Study Day End of Open-label M24 Study Period Date +**31**

[a] If results from baseline DXA are not available, the results from scan taken on or before Study Day 14 will be considered baseline values and not the Month 6 values.

13.1.1.4 QCT for imaging component of Sub-study

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline [a]	1	Last evaluation prior to or on Study Day 1
Month 6	183	Study Day 2 to 274
Month 12	366	Study Day 275 to End of Double-blind Period Date + **30**
Month 24	731	≥ End of Double-blind Period Date +**31**

[a] If results from baseline QCT is not available, the results from scan taken on or before Study Day 14 will be considered baseline values and not the Month 6 or Month 12 values.

13.1.1.5 OC and BSAP for Sub-study

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline	1	Last evaluation prior to or on Study Day 1 before dosing
Month 1	31	Study Day 2 to 61
Month 3	92	Study Day 62 to 137
Month 6	183	Study Day 138 to 229
Month 9	275	Study Day 230 to 320
Month 12	366	Study Day 321 to End of Double-blind Period Date

13.1.1.6 BTM (Excluding OC and BSAP) or Biomarker for Sub-study

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline	1	Last evaluation prior to or on Study Day 1 before dosing
Month 1	31	Study Day 2 to 61
Month 3	92	Study Day 62 to 137
Month 6	183	Study Day 138 to 229
Month 9	275	Study Day 230 to 320
Month 12	366	Study Day 321 to End of Double-blind Period Date
Month 18	548	End of Double-blind Period Date +1 to Study Day 639
Month 24	731	Study Day 640 to End of Open-label M24 Study Period Date
Month 36	1096	End of Open-label M24 Study Period Date+1 to Study Day 1278
Month 48	1461	Study Day 1279 to 1643
Month 60	1826	Study Day 1644 to 2008

Data are collected every 12 months after Month 36 until the end of the primary analysis period.

13.1.1.7 Laboratory

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline	1	Last evaluation prior to or on Study Day 1 before dosing
Month 1	31	Study Day 2 to 61
Month 3	92	Study Day 62 to 137
Month 6	183	Study Day 138 to 229
Month 9	275	Study Day 230 to 320
Month 12	366	Study Day 321 to End of Double-blind Period Date
Month 18	548	End of Double-blind Period Date +1 to 685
Month 24	731	Study Day 686 to End of Open-label M24 Study Period Date
Month 30	913	End of Open-label M24 Study Period Date +1 to Study Day 1004
Month 36	1096	Study Day 1005 to 1187
Month 42	1278	Study Day 1188 to 1369
Month 48	1461	Study Day 1370 to 1552
Month 54	1643	Study Day 1553 to 1734
Month 60	1826	Study Day 1735 to 1917

Data are collected every 6 months after Month 36 until the end of the study. To be processed and sent to the central laboratory during the primary analysis period; to be analyzed at local laboratories following the primary analysis period.

13.1.1.8 Antibody (AB)

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline	1	Last evaluation prior to or on Study Day 1
Month 1[a]	31	Study Day 2 to 61
Month 3	92	Study Day 62 to 137
Month 6	183	Study Day 138 to 274
Month 12	366	Study Day 275 to End of Double-blind Period Study Day
Month 18	548	End of Double-blind Period Date +1 to 685
Month 24	731	≥ Study Day 686

[a] Any antibody assessment done on Study Day 1 but after the administration of the investigational product will be classified into Month 1.

13.1.1.9 Weight and Height

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline	1	Last evaluation prior to or on Study Day 1
Month 6	183	Study Day 2 to 274
Month 12	366	Study Day 275 to End of Double-blind Period Date
Month 18	548	End of Double-blind Period Date +1 to Study Day 639
Month 24	731	Study Day 640 to End of Open-label M24 Study Period Date
Month 30	913	End of Open-label M24 Study Period Date +1 to Study Day 1004
Month 36	1096	Study Day 1005 to 1187
Month 42	1278	Study Day 1188 to 1369
Month 48	1461	Study Day 1370 to 1552
Month 54	1643	Study Day 1553 to 1734
Month 60	1826	Study Day 1735 to 1917

Data are collected every 6 months after Month 36 until the end of the study.

13.1.1.10 OPAQ-SV, LAD, BPI and EQ-5D

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline	1	Last evaluation prior to or on Study Day 1
Month 6	183	Study Day 2 to 274
Month 12	366	Study Day 275 to End of Double-blind Period Date
Month 18	548	End of Double-blind Period Date +1 to Study Day 639
Month 24	731	Study Day 640 to End of Open-label M24 Study Period Date
Month 30	913	End of Open-label M24 Study Period Date +1 to Study Day 1004
Month 36	1096	Study Day 1005 to 1187
Month 42	1278	Study Day 1188 to 1369
Month 48	1461	Study Day 1370 to 1552
Month 54	1643	Study Day 1553 to 1734
Month 60	1826	Study Day 1735 to 1917

Data are collected every 6 months after Month 36 until the end of the primary analysis period.

13.1.1.11 Post-fracture OPAQ-SV and EQ-5D-5L, LAD and BPI

Visit Relative to Fracture	Target Day	Post-fracture Visit Window Definition
Pre-fracture Baseline	NA	Last evaluation prior to the reported nonvertebral or clinical fracture date
Month 1 *	31	The scheduled monthly post-fracture visit within day 16 to day 45 post-fracture window
Month 2 *	61	The scheduled monthly post-fracture visit within day 46 to day 75 post-fracture window
Month 3 *	91	The scheduled monthly post-fracture visit within day 76 to day 105 post-fracture window

*If more than one actual visit falls within the same defined window, the closest visit to the target day will be considered for analysis. If two assessment actual dates are at the same distance from the target day, the later visit will be considered for analysis.

13.1.1.12 Vital Signs

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline	1	Last evaluation prior to or on Study Day 1
Month 1[a]	31	Study Day 2 to 106
Month 6	183	Study Day 107 to 274
Month 12	366	Study Day 275 to End of Double-blind Period Date
Month 18	548	End of Double-blind Period Date +1 to Study Day 639
Month 24	731	Study Day 640 to End of Open-label M24 Study Period Date
Month 30	913	End of Open-label M24 Study Period Date +1 to Study Day 1004
Month 36	1096	Study Day 1005 to 1187
Month 42	1278	Study Day 1188 to 1369
Month 48	1461	Study Day 1370 to 1552
Month 54	1643	Study Day 1553 to 1734
Month 60	1826	Study Day 1735 to 1917

[a] Any laboratory or antibody assessment done on Study Day 1 but after the administration of the investigational product will be classified into Month 1.

Data are collected every 6 months after Month 36 until the end of the primary analysis period.

13.2 Genant Semiquantitative Scoring Method

Both incident and prevalent vertebral fractures present on a radiograph are defined by using the Genant Semiquantitative Scoring Method (Genant et al, 1993) for each vertebra (T4 through L4) as described in the table below.

Grade	Fracture Severity	Definition
0	Normal	Normal; (Approximately less than 20% reduction in anterior, middle, and/or posterior height)
1	Mild	Approximately 20% to 25% reduction in anterior, middle, and/or posterior height
2	Moderate	Approximately 25% to 40% reduction in anterior, middle, and/or posterior height
3	Severe	Approximately 40% or greater reduction in anterior, middle, and/or posterior height

13.3 Reliability, Validity, and Scoring Algorithm for PRO Measurements

13.3.1 OPAQ-SV

Reliability and Validity

The OPAQ-SV is a comprehensive instrument designed to evaluate the impact of vertebral and nonvertebral fractures on different aspects of health-related quality of life. The OPAQ-SV contains 34 questions that can be summarized into 7 scales (walking/bending, daily activities, transfer, fear of falls, back pain, body image, and independence) and 3 dimensions (physical function, emotional status, and back pain). The OPAQ-SV has been validated in osteoporotic subjects and used extensively. Internal consistency for the OPAQ-SV is high (Cronbach's alpha = 0.86; Silverman SL, 2000).

Question Mapping

Dimension	Scale	Questions
Physical function	Walking/bending	1 – 7
	Daily activities	8 – 15
	Transfer	16 – 19
Emotional status	Fear of falls	20 – 24
	Body image	29 – 31
	Independence	32 – 34
Back pain	Back pain	25 – 28

Scoring Algorithm

1. Standardize the response for each question to a 0 – 100 scale according to the table below.
2. Average the standardized values within each of the 7 scales and 3 dimensions to create the summary or scale scores based on the question mapping. The score will only be calculated if ≥ 50% of the items within the scale or dimension are nonmissing. Otherwise, set the score to missing.

Question Number	Question Response	Question Score	Standardized Value
2, 5, 6, 7, 16 - 19, 25, 27, 28	All days	1	0
	Most days	2	25
	Some days	3	50
	Few days	4	75
	No days	5	100
1, 3, 4, 8, 9, 10	All days	1	100
	Most days	2	75
	Some days	3	50
	Few days	4	25
	No days	5	0
11,12, 20,21, 22, 23, 24, 29, 30, 31, 33, 34	Always	1	0
	Very often	2	25
	Sometimes	3	50
	Almost never	4	75
	Never	5	100
13, 14, 15, 32	Always	1	100
	Very often	2	75
	Sometimes	3	50
	Almost never	4	25
	Never	5	0
26	Severe	1	0
	Moderate	2	25
	Mild	3	50
	Very mild	4	75
	None	5	100

13.3.2 EQ-5D-5L

Visual Analog Scale

Questionnaire asks respondents to rate their present health status on a vertical 0 to 100 visual analog scale of 20 cm, with 0 labeled as “Worst imaginable health state” and 100 labeled as “Best imaginable health state.” The scale is marked in increments of “1,” with values labeled at each decile. The person scoring this question must observe

the point at which the subject's hand-drawn line intersects the scale and enter the closest integer.

13.3.3 LAD Questionnaire

Limited activity days questionnaire evaluate subjects' back pain experience during the past 30 days and this questionnaire has been used in several clinical trials including the Fracture Intervention Trial (Nevitt MC, 2000). The responses of each question will be analyzed separately. The responses of each question are as follows:

Questions	Responses
1, 2a, 2b	If question 1 = 'Yes', then answer question 2a and 2b. 2a = Specify number of days cutting down on things due to health reasons or mark 'Don't know' 2b = Specify number of days cutting down on things due to a fracture If question 1 = 'No' or 'Subject does not know', then answer question 3a.
3a, 3b	If question 3a = 'Yes', then specify number of nights spent in a hospital due to health reasons and then answer question 3b. 3b = Specify number of nights spent in a hospital due to a fracture. If question 3a = 'No' or 'Subject does not know', then answer question 4a.
4a, 4b	If question 4a = 'Yes', then specify number of days staying in bed for more than half the day (NOT in a hospital) more than half of the day due to health reasons and then answer question 4b. 4b = Specify number of days spent in bed for more than half the day due to a fracture. If question 4a = 'No' or 'Subject does not know', then questionnaire is complete.

The responses of each question will be analyzed individually.

13.4 Code Fragments

CCI

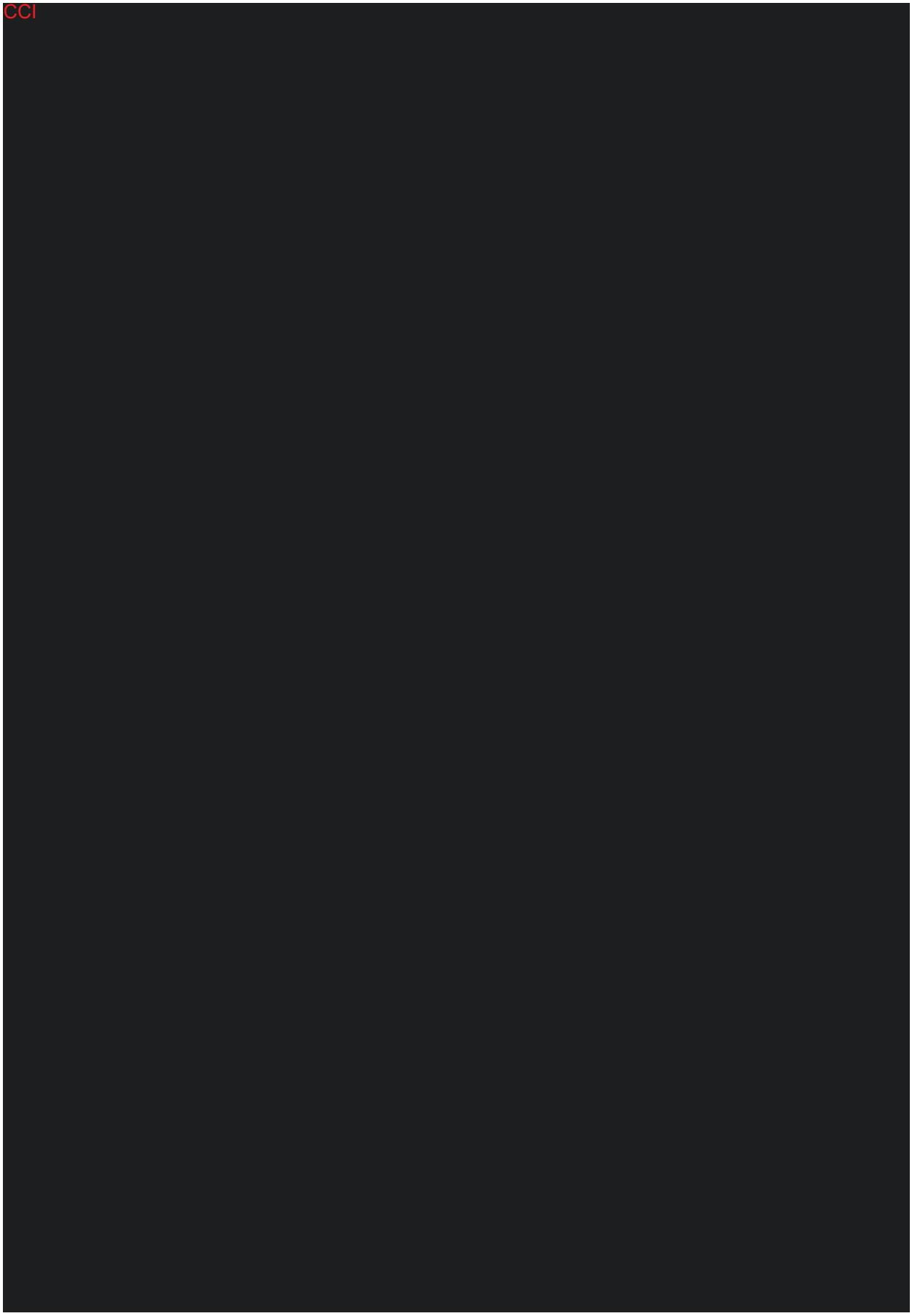
CCI

[redacted]

CCI

[redacted]

CCI



CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

13.5 Definitions of Adjusted *p*-values

Hochberg procedure will be applied to adjust the p-values for the two primary endpoints, new vertebral fracture through M24 and clinical fracture through PA. For illustration purpose, the p-value for new vertebral fracture through M24 is assumed to be the smaller of the 2 p-values. The adjusted p-values will be compared to 0.05 to determine whether the endpoint is significant except the nonvertebral fracture through PA will be compared to the α that is determined by the alpha spending function.

Test ID	Endpoint	Nominal p-value	Adjusted p-value
Two primary endpoints			
1	New vertebral fracture through M24	p_1	Min (p_1* 2, p_2)
2	Clinical Fracture through PA	p_2	p_2
Secondary endpoints specified in the testing sequence			
3	Lumbar spine at Month 24	p_3	Max (p_2, p_3)
4	Total Hip at Month 24	p_4	Max (p_2, p_3, p_4)
5	Femoral neck at Month 24	p_5	Max (p_2, p_3, p_4, p_5)
6	Lumbar spine at Month 12	p_6	Max (p_2, p_3, p_4, p_5, p_6)
7	Total Hip at Month 12	p_7	Max (p_2, p_3, p_4, p_5, p_6, p_7)
8	Femoral neck at Month 12	p_8	Max (p_2, p_3, p_4, p_5, p_6, p_7, p_8)
9	Nonvertebral fracture through PA (1-sided) & information fraction = I	p_9	Max (p_2, p_3, p_4, p_5, p_6, p_7, p_8, a* 2), where (a * ln(1+(e-1) * I)) = p_9